UNIVERSITE PARIS-SUD - FACULTE DE MEDECINE PARIS SUD

THESE POUR LE DOCTORAT EN MEDECINE

31

ANNÉE

(Diplôme d'Etat)
par

N

NOM et Prénoms du Candidat : RISSE René, Albert, Jean
Né le : 18 Juin 1932 à PARIS (14ème)

PRESENTEE et SOUTENUE PUBLIQUEMENT LE : 1972

TITRE DU MANUSCRIT

REFLEXION DOCIMOLOGIQUE
A propos de deux examens de deuxième année du Premier cycle des Etudes MEDICALES.

Le Président : Monsieur MONSAINGEON, Doyen.

UNIVERSITE PARIS-SUD - FACULTE DE MEDECINE PARIS SUD

Année

N°

THESE POUR LE DOCTORAT EN MEDECINE

(Diplôme d'Etat)
par

NOM et Prénoms du Candidat : RISSE René, Albert, Jean
Né le : 18 Juin 1932 à PARIS (14ème)

PRESENTEE et SOUTENUE PUBLIQUEMENT LE :

TITRE DU MANUSCRIT

REFLEXION DOCIMOLOGIQUE

A propos de deux examens de deuxième année du Premier cycle des Etudes MEDICALES.

Le Président : Monsieur MONSAINGEON, Doyen.

UNIV.
LEIDEN
BIBL.

A Monsieur BARRES, Professeur de Physiologie.

Il y a peut être d'autres Professeurs que vous,Monsieur, qui, sans faiblesse aucune, acceptent que leur enseignement soit critiqué et la valeur de leurs examens mise en doute mais,c'est un fait, je n'en connais pas d'autres.

Physiologiste, vous n'avez pas oublié, en devenant enseignant, ce que toute mesure exige de rigueur, toute hypothèse de réflexion, toute expérimentation de prudence ni tout enseignement de responsabilité. En proclamer l'exemplarité, c'est de la reconnaissance.

Vous m'avez encouragé dans une voie que j'osais à peine prendre, en veuillant bien déclarer qu'un tel travail pouvait être intéressant. Vous m'avez aidé pas à pas et parfois stimulé, par exemple, en me donnant un très précieux moyen d'établir les résultats; à ce sujet, parler de démagogie lorsqu'on connaît l'auteur fait sourire. La confiance que vous avez bien voulu me manifester, l'accueil que vous m'avez offert, non en me disant mais en me répétant que, dans votre service, j'étais chez moi, tout cela, Monsieur,appelle une gratitude que je vous demande de croire sincère.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549217_0032

R E M E R C I E M E N T S,

A mes parents :

à mon père qui, bien avant l'Université,
m'a appris à apprendre ;

à ma mère qui m'a aidé à entendre au-delà des mots .....
et qui n'ont pu réussir que dans la limite de mes propres moyens.

A ma femme,

qui a découvert que certaines gestations durent plus de neuf mois et qui m'a apporté une aide dont je suis seul à pouvoir connaître le prix.

-:-:-:-

Cinquante et un pour cent de reçus, c'est une manière de résumer des résultats d'examen. Quarante neuf pour cent de refusés, c'est une autre manière qui ouvre d'autres pespectives. Deux cent soixante seize étudiants sur trois cent trente deux sont devenus candidats: pourquoi ? Un rendement de 50 p.100 est-il bon, est-il mauvais, est-il au moins significatif? De quoi ? D'un mauvais niveau des étudiants, d'un mauvais enseignement ou d'un examen mal fait ?

A un certain moment, même un enseignant éprouve le besoin de comprendre le sens de son action, son efficacité, la nature de ses qualités ou de ses défauts et la découverte du fait que presque 54 p.100 de la population prise en charge échoue finalement entraîne une réflexion.

Des réflexes surgissent: l'analyse de l'examen en ses cinq épreuves constitutives amène à comparer les effectifs pour chacune d'elles: ils sont différents. Les moyennes sont les mêmes, ou à peu près, quels rapports unissent ces deux grandeurs ? L'échec dans la tentative de réponse à cette question entraîne le franchissement d'un pas: étudions les variances et les corrélations. L'impossibilité persistante de répondre au genre de question " celui qui reçoit 80 p.100 des candidats est-il un horrible démagogue et celui qui ne reçoit que 40 p.100 est-il un abominable sadique ?" amène à s'interroger sur la valeur des renseignements traditionnellement retenus comme fondamentaux et la réflexion commence. Elle aboutit à une constatation déprimante: même l'utilisation de savantes techniques de statistiques dans lesquelles on s'est engagé à reculons et comme malgré soi ne fournit rien d'intéressant. A l'époque où les musiciens ne connaissaient pas le diapason (jusqu'au milieu du XIXe siècle) ils savaient mieux s'accorder que les examinateurs car s'ils ne connaissaient pas de référence universelle, ils utilisaient tous les mêmes écarts.

Inéluctablement, ce genre de réflexion devait amener à constater ce que certains prophètes prêchent dans le désert depuis le début du XXè siècle: le système des examens est une forme sans fond. Une réflexion docimologique ne pouvait qu'aboutir à une conclusion déprimante car négative. Pourquoi donc, après avoir contribué à la tentative de destruction d'un édifice utile, certes mais mal conçu, essayer de proposer des solutions de remplacement aux fins de construire après avoir démoli ? Pour ne pas exprimer une critique purement négative et, après avoir dénoncé un mal, proposer quelques éléments de son remède.

CHAPITRE I

EFFECTIFS

L'effectif des candidats déclarés satisfaisants dans chaque discipline est le nombre le plus immédiatement accessible et, sans doute pour cette raison, retenu comme caractéristique. Lorsque cinq épreuves constituent un examen, les comparaisons des cinq effectifs élémentaires (d'épreuves) entre eux, d'une part et avec l'effectif global (de l'examen), d'autre part, ménagent quelques surprises.

- VALEUR DES EFFECTIFS 3
- COMPARAISON DES EFFECTIFS 4
- SERIES: HYPOTHESE D'EPREUVES SUCCESSIVES II
- COMBINAISONS: SERIES APPARIEES I5
- TABLEAU 7: ROLE DES COMPENSATIONS I8
- SYNTHESE: RESULTATS BINAIRES 23
- EVOLUTION: SITUATION EN I97I 25

  Ière hypothèse: toutes les disciplines reçoivent le même effectif (26); 2ème hypothèse: les non reçus dans une discipline sont joués à pile ou face dans les autres (27); 3ème hypothèse: l'examen sélectionne par pile ou face (28).

- CONCLUSIONS : Intérêt de l'étude des notes 29

N.B.- Les pages numérotées 30 et 3I n'existent pas.

V A L E U R S

Un examen est passé. Deux cent soixante seize candidats ont subi les cinq épreuves qui le composaient. Ont obtenu au moins dix sur vingt (la "moyenne") :

En A	207
En BC	171
En BP	156
En H	200
En PHY	91

Ces valeurs correspondent à celles que fournissent les histogrammes qui seront utilisés ultérieurement. Une constatation s'impose : les effectifs reçus dans chaque discipline seront différents. L'application de deux écarts types en plus ou en moins de l'effectif observé donne :

	$- 2 \varepsilon$	effectifs reçus (n)	$+ 2 \varepsilon$	effectifs reçus p.100
A	193	207	221	0,75
BC	155	171	187	0,62
BP	140	156	172	0,57
H	185	200	215	0,72
PHY	76	91	106	0,31

Les nombres exprimant des effectifs sont arrondis à l'unité la plus proche.

Les"pourcentages" (en réalité : fraction) de l'effectif reçu sont exprimés en valeurs arrondies à la seconde décimale la plus proche.

C O M P A R A I S O N S

Que ces valeurs soient différentes est une évidence, mais une question se pose : les écarts entre les pourcentages observés sont-ils significatifs ? Pour y répondre, on pourra comparer entre eux tous les couples possibles, c'est-à-dire dix. On établira l'écart réduit de leur différence :

$\varepsilon < 1{,}96$ permettant de dire que l'écart n'est pas significatif.
$\varepsilon > 1{,}96$ permettant d'affirmer le caractère significatif de l'écart et d'apprécier la grandeur du risque (α)

Soient deux pourcentages à comparer :

pX celui de la discipline X, et
pY celui de la discipline Y

l'effectif (n) étant, ici, le même pour X et pour Y, on a :

$$\varepsilon = \frac{|pX - pY|}{\sqrt{\frac{pX\,(1-pX)}{nX} + \frac{pY\,(1-pY)}{nY}}}$$

Par exemple :

en A : $pA = 0{,}75$; $(1 - pA) = (1-0{,}75) = 0{,}25$; $n = 276$
en BC : $pBC = 0{,}62$; $(1-pBC) = (1-0{,}62) = 0{,}38$; $n = 276$

$$pA.pBC = \frac{pA - pBC}{\sqrt{\frac{pA\,(1-pA)}{276} + \frac{pBC\,(1-pBC)}{276}}} =$$

$$\frac{0{,}75 - 0{,}62}{\sqrt{\frac{0{,}75 \times 0{,}25 + 0{,}62 \times 0{,}38}{276}}} =$$

$$\frac{0{,}13}{0{,}039} = 3{,}33$$

ce qui permet de retenir l'écart comme significatif au risque inférieur à 0,001 ou 10^{-3}

Pour l'ensemble des disciplines les valeurs trouvées sont les suivantes (le risque α correspondant est indiqué si la différence est significative).

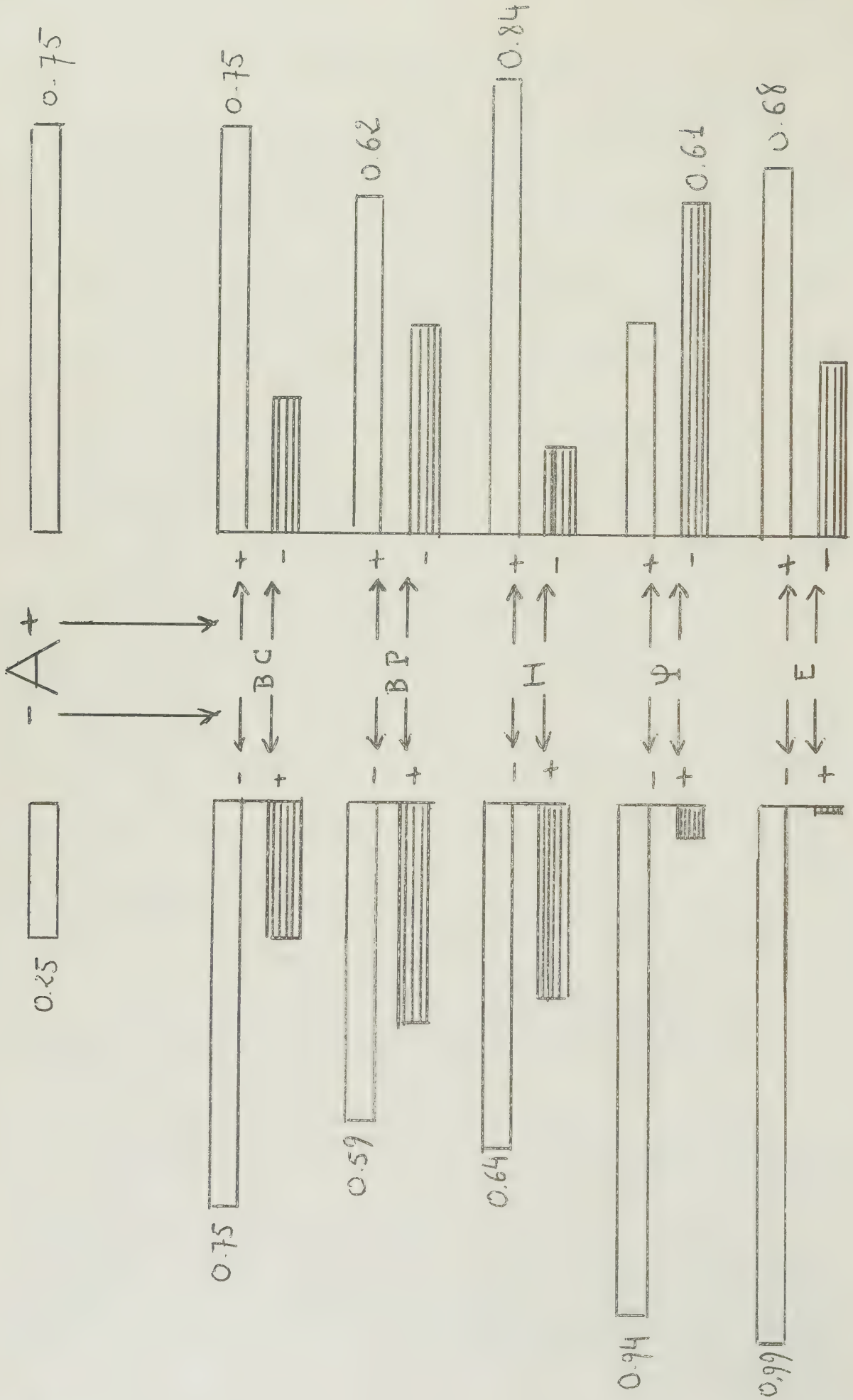
Comparaison des p.100	Valeur de ε	Valeur de α
A.BC	3,33	$< 10^{-3}$
A.BP	4,61	$< 10^{-5}$
A.H	0,81	
A.PHY	11,50	$< 10^{-9}$
BC.BP	1,21	
BC.H	2,56	< 0,02
BC.PHY	7,75	$< 10^{-9}$
BP.H	3,75	$< 10^{-3}$
BP.PHY	6,50	$< 10^{-9}$
H.PHY	10,78	$< 10^{-9}$

Deux seulement des dix écarts observés ne sont pas significatifs. L'un est à la limite. On peut donc retenir :

. Il existe une liaison entre A et H
. Il existe une liaison entre BC et BP
. Un certain rapprochement peut se faire entre BC et H
. Entre PHY et toutes les autres disciplines l'écart est hautement significatif.

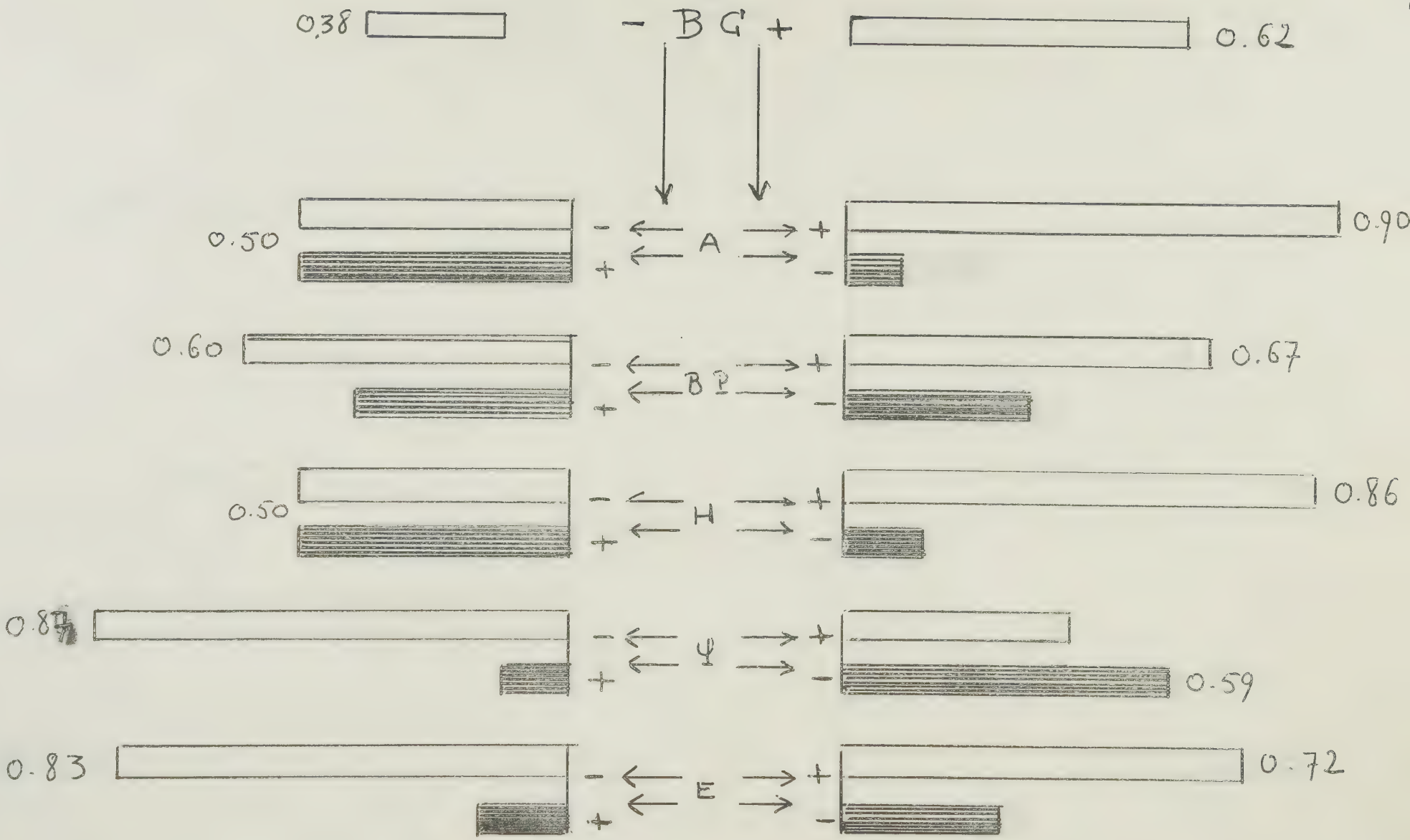
Cependant, il n'est pas inutile de ranger ces disciplines en fonction de la valeur du risque qui s'attache au pari sur le caractère significatif des écarts : BP, BC, H et A (ε valant respectivement : 6,5 - 7,75 - 10,78 - 11,5).

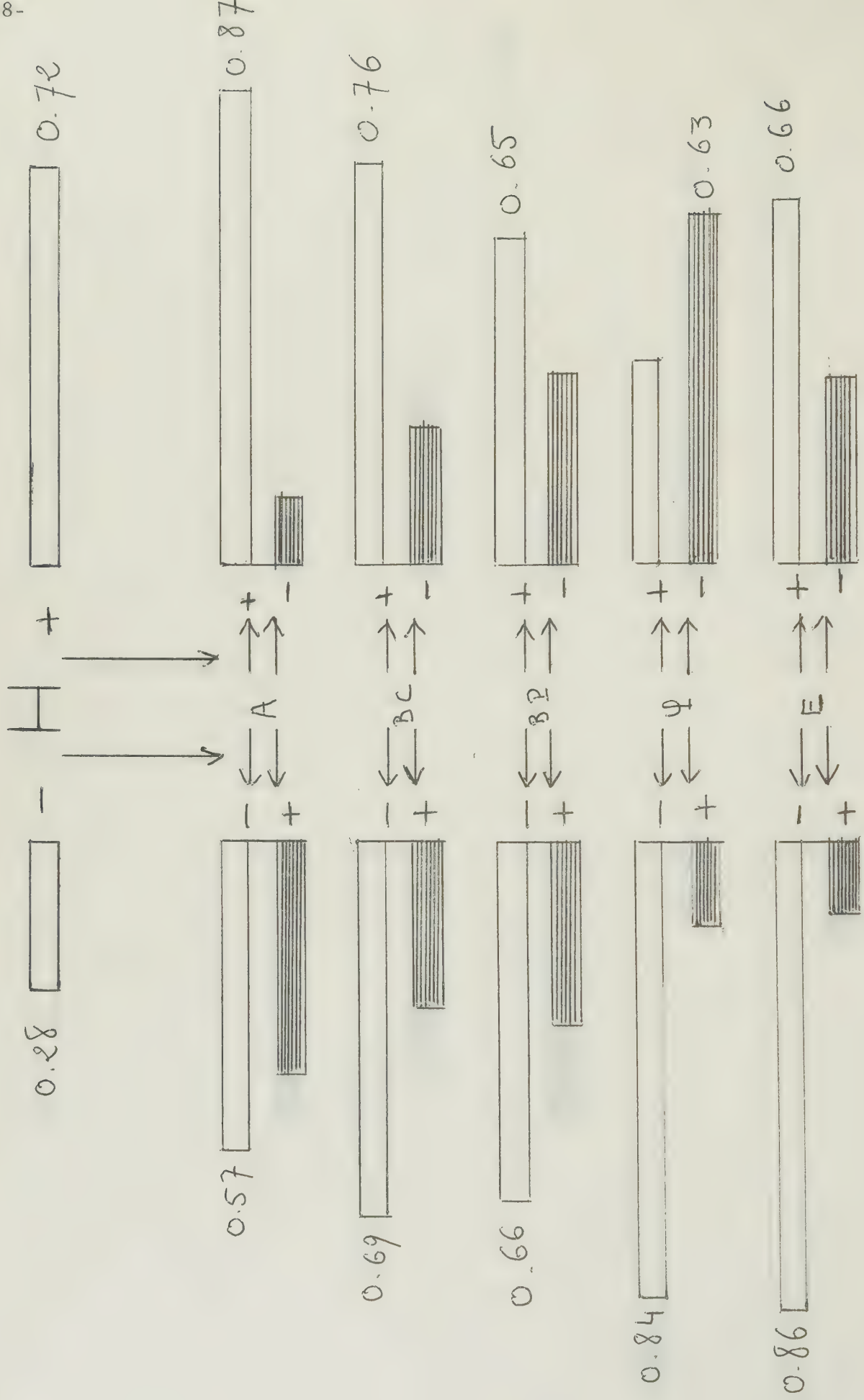
Les valeurs ainsi trouvées correspondent aussi bien à la comparaison des pourcentages des reçus qu'à celle des non reçus. En effet, dans le rapport permettant le calcul de "ε" le dénominateur ne changerait pas, qu'on prenne en considération la fraction de réussite ou celle des échecs. Le numérateur ne changerait pas non plus : au lieu de l'exprimer par pX - pY, on l'exprimerait par (1 - pX) - (1 - pY) ce qui revient exactement au même.



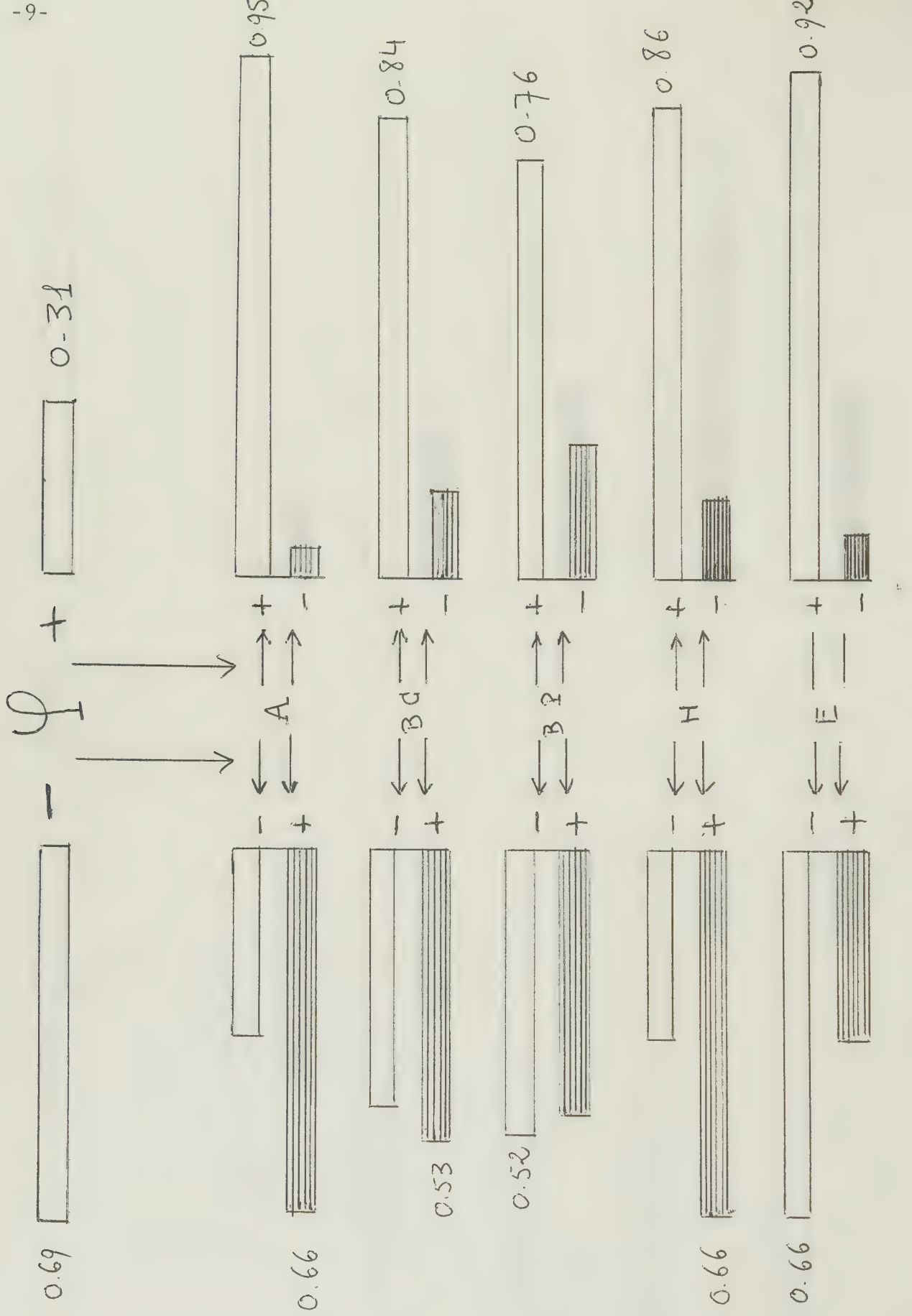
GRAPHIQUE N° 1: entrée A.

GRAPHIQUE N° 2:entrée BC.

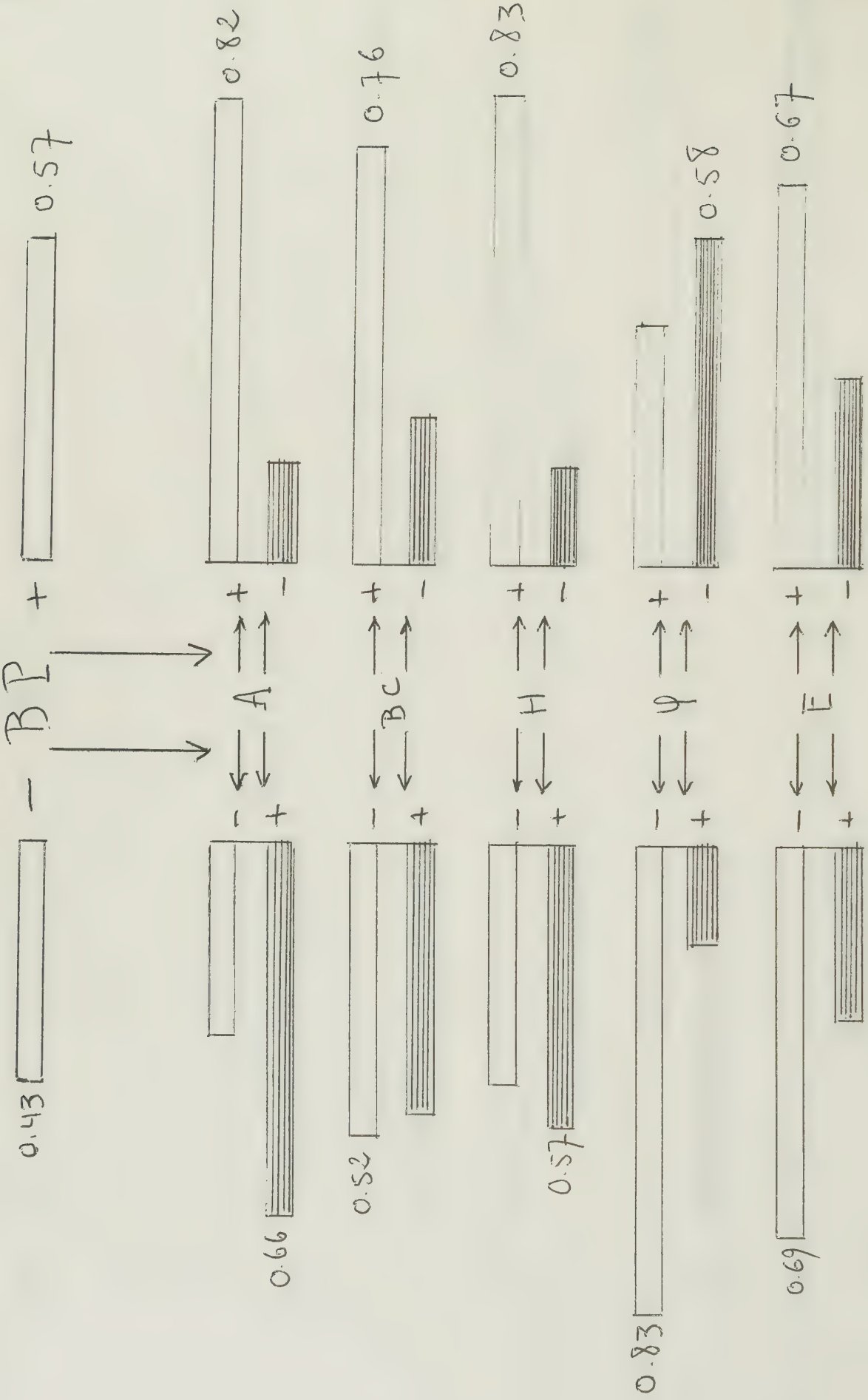




GRAPHIQUE N° 3:entrée H.



GRAPHIQUE N° 4:entrée PHY.



GRAPHIQUE N° 5:entrée BP.

S E R I E S

Si la comparaison des pourcentages, telle qu'elle a été faite, montre l'hétérogénéité des disciplines quant à l'effectif des candidats qu'elles reçoivent, elle suggère fallacieusement que ces disciplines reçoivent ou refusent de façon homogène lorsqu'on les examine deux par deux.

Soit une discipline X présentant ou non, par rapport à une discipline Y, une différence significative pX - pY. On peut s'attendre à ce qu'elles soient en accord (ou en désaccord égal) pour recevoir ou refuser. Il est intéressant de vérifier s'il en est bien ainsi.

Si : pX^+ et pY^+ représentent les pourcentages de reçus en X et Y
pX^- et pY^- celui des refusés

on se pose la question : "Si on imagine que les candidats passent successivement X et Y ou Y et X, que deviennent-ils en Y selon le sens de leurs résultats en X ?"

Pour la séquence XY, les candidats sont d'abord jugés en X : ils deviennent X^+ ou X^-. Les X^+ peuvent devenir Y^+ ou Y^-. De même les X^- : ils peuvent devenir Y^+ ou Y^-. Si on compare en X^+, pY^+ et pY^-, par établissement de l'écart réduit de leur différence, il deviendra possible de comparer entre eux tous les couples d'après la valeur de ε.

$$\varepsilon = \frac{pY^+ - pY^-}{\sqrt{\frac{pY^+ \times pY^-}{n} + \frac{pY^+ \times pY^-}{n}}}$$

où n représente : soit l'effectif réel des X^+,
soit celui des X^-.
On pourrait de même comparer entre eux les X^+ et les X^- à partir des Y^+ ou des Y^-.

La comparaison des cinq graphiques : 1, 2 , 3, 4 et 5 montre l'intérêt d'une telle opération. Par exemple, le graphique BP : il montre ce que deviennent les "candidats BP" lorsqu'ils sont soumis à un second contrôle.
- A droite, ceux qui sont reçus (BP^+ : 156)
- A gauche, ceux qui ne le sont pas (BP^- : 120)

De chaque côté, non pas dix rangées, mais cinq paires de rangées.

- En haut et à droite, la première paire représente que, sur 156 candidats BP^+ (100 p.100 de BP^+), 128 sont A^+ (82 p.100) et 28 sont A^- (18 p.100).
- En haut et à gauche, la paire la plus haute représente l'avenir en A des 120 candidats BP^-. Ils se répartissent en 41 A^- (34 p.100) et 79 A^+ (65 p.100).

Il apparaît ainsi que les candidats refusés en BP ont plus de chances d'être reçus (65 p.100) que refusés (34 p.100) en A !

TABLEAU N° 6

RECUS EN	A^+	BC^+	BP^+	H^+	PHY^+
A^+_-		25	14 88	22 40	27 20
BC^+_-	11 90		9 79	11 10	12 14
BP^+_-	5 10	6 80		6 38	8 00
H^+_-	18 00	19 00	15 70		12 83
PHY^+_-	- 4 68	- 3 39	- 2 90	- 5 41	

NNON RECUS:	A^-	BC^-	BP^-	H^-	phy^-
A^+_-		0 00	- 5 24	1 77	- 6 66
BC^+_-	3 45		0 62	5 13	- 1 17
BP^+_-	2 16	2 98		4 21	0 78
H^+_-	3 45	0 00	- 2 18		- 6 66
PHY^+_-	22	16	13 75	11 50	

Comparaison des effectifs (en p.100)
Les valeurs sont celles des écarts réduits.

Un rapide coup d'oeil sur les cinq graphiques montre qu'à droite comme à gauche, les rapports peuvent être plutôt concordants (colonne claire plus grande) ou plutôt discordants (colonne sombre plus grande) ou indifférents (colonne claire et sombre d'égales longueurs). C'est cette impression que le calcul veut confirmer.

Exemple sur BP^-, $A^\pm$:

Nombre de refusés en BP : $nBP^- = 120$

$pA^- = 0,34$

$pA^+ = 0,66$

$$\varepsilon = \frac{|0,34 - 0,66|}{\sqrt{\frac{0,34 \times 0,66}{120} + \frac{0,34 \times 0,66}{120}}} =$$

$$\frac{|0,34 - 0,66|}{\sqrt{\frac{0,34 \times 0,66 \times 2}{120}}} = - 5,24$$

La grandeur ($\varepsilon > 1,96$) indique que la différence est significative. Le signe (-) indique que l'accord est discordant : les"mauvais" en BP sont plutôt "bons" en A.

Les résultats concernant les vingt couples sont groupés dans le tableau 6. (Provisoirement les résultats globaux de l'examen n'ont pas été pris en considération.)

Le tableau comporte cinq colonnes, correspondant chacune à une épreuve considérée comme initiale. Chaque colonne est divisée en deux moitiés :
- l'une supérieure, concerne l'avenir de ceux qui ont été acceptés dans la discipline
- l'autre inférieure, concerne l'avenir de ceux qui ont été refusés.

Dix lignes représentent l'avenir des candidats pour chaque discipline considérée comme secondaire.

Les cases comportent les valeurs de "ε" sans signe si les résultats sont "concordants" et l'écart significatif.

"ε" est précédé est précédé du signe - si les résultats sont discordants.

Si "ε" n'est pas l'expression d'un écart significatif, sa valeur est encadrée.

Plusieurs remarques semblent s'imposer :

1°) A et H sont en accord pour accepter tous ceux qui l'ont déjà été : on pourrait dire avec enthousiasme ($\varepsilon \geqslant 12,83$).
2°) BC et BP manifestent le même accord mais avec moins d'enthousiasme ($\varepsilon \leqslant 12,14$).
3°) PHY est plutôt en désaccord pour recevoir.
4°) Les associations AH et BC-BP avaient déjà été remarquées : ces deux couples ne présentaient pas d'écart significatif quant aux pourcentages de leurs acceptés.

5°) Si on excepte PHY, les 16 couples restant sont en accord pour accepter. En incluant PHY, 16 couples sur 20 sont significativement en accord.

6°) Ce bel accord s'effrite lorsqu'il s'agit de refuser :

a) Si on excepte PHY, sur les 16 couples restant, 6 seulement sont en accord, 4 sont en désaccord significatif et 6 sont hésitants dont 2 parfaitement ($\varepsilon = 0$). En incluant PHY, on ajoute 4 couples d'accord (donc au total, 10 sur 20 seulement).

b) L'association A-H manque d'harmonie : H confirme (mollement) les refus de A, mais A ne se prononce pas sur ceux de H. Tous deux restent en parfait accord pour ne pas se prononcer vis-à-vis de BC et pour infirmer les résultats de PHY. Quant à l'association BC-BP elle ne se prononce pas très fort ($\varepsilon \leqslant 5{,}13$). Elle ne se prononce pas sur le jugement de PHY, BC ne confirme, ni n'infirme, le résultat de BP tandis que BP approuve BC.

c) PHY est franchement de l'avis de refuser avec force ($\varepsilon \geqslant 11{,}5$) mais avec une décroissance qui ne rappelle pas du tout la précédente : l'ancien ordre : BP - BC - H - A
est devenu : A - BC - BP - H
Les deux disciplines les moins en accord quant à la proportion de reçus (A et PHY) manifestent le plus grand accord pour refuser.

Au total, l'accord est bien meilleur pour recevoir que pour refuser.

C O M B I N A I S O N S

Les dix couples envisagés ont été étudiés jusqu'ici en ignorant que les différentes épreuves avaient été subies par les mêmes candidats au risque de conclure, à tort, au caractère non significatif de l'écart des pourcentages observés. On peut réparer cette éventuelle erreur en comparant les séries appariées : A.BC . A.BP ..... etc. Chaque série se représente ainsi :

	Y	
II		III
$X^- Y^+$		X^+Y^+
		X
$X^- Y^-$		X^+Y^-
I		IV

Les quadrants impairs indiquent des accords : (--) ou (++) tandis que les quadrants pairs indiquent les désaccords (+-) ou (-+). La question est de savoir si les désaccords sont significatifs. Pour cela on calcule l'écart réduit "ε" ainsi :

$$\varepsilon = \frac{|n(X^+Y^-) - n(X^-Y^+)|}{\sqrt{n(X^+Y^-) + n(X^-Y^+)}}$$

Exemple :

	A	
52		155
		BC
52		17

$$\varepsilon = \frac{52 - 17}{\sqrt{52 + 17}} = \frac{35}{\sqrt{69}} = 4{,}213 \text{ (significatif pour } \alpha < 10^{-4})$$

Les résultats fournis par le tableau suivant sont comparables à ceux qu'avait fourni l'étude des différences des pourcentages observés.

couples	ε	α
A - BC	4,213	10^{-4}
A - BP	4,930	$< 10^{-6}$
A - H	1,050	
A	10,700	$< 10^{-9}$
BC - BP	1,616	
BC - H	3,117	$< 10^{-2}$
BC	8,112	$< 10^{-9}$
BP - H	4,411	$\simeq 10^{-5}$
BP	6,739	$< 10^{-9}$
H	9,704	$< 10^{-9}$

Ce qui a été vu précédemment peut ainsi se résumer : supposons que les candidats passent successivement deux épreuves ; par exemple A puis BC. La différence observée entre ceux qui avaient été initialement acceptés en A, puis secondairement refusés ou acceptés en BC est-elle significative ? En d'autres termes, a-t-on plus de chance d'être accepté en Y si on est accepté en X que si on y est refusé ? L'action réciproque des résultats d'une discipline sur ceux d'une autre a fait apparaître un nombre relativement élevé de contradictions surtout si on considère l'action de refuser.

Il est possible de refaire le tableau 6 en modifiant son entrée : la question qu'on pourra alors se poser sera:la différence des pourcentages est-elle encore significative lorsqu'on exige non plus une condition préalable mais deux, la deuxième étant la réussite à l'examen ? L'examen du couple A.BC en détail permettra de démonter le mécanisme. En A, 207 candidats ont été jugés satisfaisants. En BC, 172. La différence entre ces deux résultats avait été jugée significative.

Parmi les 207 acceptés en A, 155 l'ont été en BC et 52 refusés. BC a confirmé les résultats de A : la différence est hautement significative ($\varepsilon = 11,9$). Inversement, parmi les 172 acceptés en BC, A a con-

firmé 155 fois et contredit 17 fois. La différence est hautement significative ($\varepsilon = 25$) ce qui confirme cette impression : lorsqu'on est reçu en BC on est reçu en A ; la différence observée ne permet pas de rejeter cette hypothèse. Il en serait tout autrement si, parmi l'effectif des reçus en A, 50 p.100 seulement était reçus en BC.

De même, si on est refusé en A, on est plutôt refusé en BC ($\varepsilon = 3,45$). Inversement, si on est refusé en BC, on a autant de chance d'être accepté que refusé en A ($\varepsilon = 0$).

Au total : la connaissance de la situation d' "accepté" en A ou en BC permet de pronostiquer le résultat de l'autre discipline ; si on sait qu'un candidat est refusé en A, on peut parier qu'il sera refusé en BC, tandis que l'incertitude totale découle de la connaissance de la situation : "refusé en BC" quant au pronostic en A.

Si, maintenant, on exige deux conditions : être reçu (ou refusé) en A et à l'examen ou en BC et à l'examen les pourcentages deviennent :

$A^+ E^+$: 0,68 - 68 p.100 des A^+ sont E^+
$BC^+ E^+$: 0,72 - 72 p.100 des BC^+sont E^+

La différence entre $A^+ E^+$ et $BC^+ E^+$ n'est plus significative ($\varepsilon = 0,85$) : la sélection des reçus en A (sélection faite par l'examen) est reçue en BC et réciproquement. En revanche, les refusés sont dans une situation très différente :

$A^- E^-$: 0,99
$BC^- E^-$: 0,83

La différence entre les $A^- E^-$ est $BC^- E^-$ est très significative ($\varepsilon = 4,44$) on ne peut raisonnablement parier sur la situation dans une discipline quand on connaît la situation dans une autre.

Le tableau 7, concernant tous les couples possibles, établi pour les conditions précédemment définies, montre que :

a) Toutes les disciplines sont en accord pour accepter ($\varepsilon \leqslant 1,27$) sauf PHY ($4,44 \leqslant \varepsilon \leqslant 5,909$). Or, ce qui les a mises en accord c'est l'examen, et l'examen, c'est la somme de toutes les disciplines. Une somme de désaccords aboutit à un accord ! Seule PHY manifeste son désaccord persistant.

b) Deux couples seulement sont en accord pour refuser : BC.H et B.P. qui, pourtant, jusqu'ici, ne s'étaient guère montrés coopérants.

L'examen n'apparaît-il pas être une somme d'incertitudes ?

TABLEAU N° 7

pX - pY	$pX^+ - pY^+$	$pX^+ \quad pY^+$ / E^+	$pX^- \quad pY^-$ / E^-	E^+ / $pX^+ \quad pY^+$	E^- / $pX^- \quad pY^-$
A BC	3 33	- 0 85	4 44	41	- 2 36
A BP	4 61	0 20	8 23	6 75	- 1 96
A H	0 81	0 43	3 17	25	0 35
A PHY	11 50	- 5 58	9 11	10 47	- 10 20
BC BP	1 21	1	2 50	2 82	0 37
BC H	- 2 56	1 27	- 0 55	- 1 38	2 72
BC PHY	7 75	- 4 44	3 40	6 40	- 7 38
BP H	- 3 75	0 20	- 2 98	- 4 18	2 32
BP PHY	6 50	- 5 31	0 55	3 45	- 7 67
H PHY	10 78	- 5 90	3 84	7 87	- 10 69

TABLEAU 7

La considération attentive du tableau 7 permet quelques constatations importantes.

Comment le lire ?

Les valeurs représentées dans les cases sont celles de ε : valeurs encadrées si elles ne sont pas significatives.

En haut de chaque colonne figure une légende très résumée et toujours bâtie selon le même modèle : pX - pY, c'est-à-dire que les valeurs de ε sont tirées de la comparaison des pourcentages de X et de Y. X varie d'une ligne à l'autre, non seulement dans sa valeur, mais dans sa nature : X de la première ligne représente A ; dans la dernière ligne : H. De même pour Y. Les symboles de disciplines sont alignés sous pX ou pY dans la colonne de gauche. L'intérêt c'est que, d'une colonne à l'autre, la soustraction se fait toujours dans le même sens. Ainsi, dans la première rangée on a toujours pA - pBC. C'est la raison qui a fait figurer un signe (-) qui montre toujours que X < Y et ce n'est pas sans importance.

Les légendes en tête des colonnes se lisent ainsi :

1ère colonne : $pX^+ - pY^+$

C'est la comparaison des pourcentages de réussite entre les deux disciplines de la rangée. Ces valeurs ont déjà été trouvées et sont simplement rappelées.

2ème colonne : $pX^+ - pY^+ \quad (E^+)$

Indique que les pourcentages comparés sont ceux des réussites à la discipline et à l'examen : ces pourcentages sont fournis par la deuxième paire de rangées de chacun des graphiques 1, 2, 3, 4 et 5.

3ème colonne : $pX^- - pY^- \quad (E^-)$

Les pourcentages sont ceux des candidats refusés à la discipline et à l'examen.

Dans les deux cas précédents, on suppose que la séquence est : disciplines puis examen : ce dernier contrôlant les résultats des précédentes. La séquence sera exactement l'inverse en 4° et 5° : ce sont les disciplines qui contrôlent les résultats de l'examen.

4ème colonne : renseignements - du graphique n° 8

Parmi les candidats reçus à l'examen (séries de droite) on compare les pourcentages de reçus discipline par discipline, tandis que en

5ème colone :

Ce sont les pourcentages des refusés qui sont comparés.

Ce long préambule permet de simplifier beaucoup l'analyse. Si on compare A et BC (1ère rangée) on constate :

1°) que A reçoit significativement plus de candidats que BC. C'est un fait. On peut supposer que A fournit plus de reçus à l'examen. Or ,

2°) A et BC fournissent la même proportion de reçus. Puisque A reçoit plus que BC, on est en droit de prévoir que A fournira moins de refusés que BC. Or,

3°) Il y a significativement plus de refusés à l'examen parmi les refusés de A que parmi ceux de BC. On peut en déduire que A est en meilleur accord avec l'examen que BC, ce qui devra se manifester par un plus fort pourcentage de reçus en A qu'en BC parmi ceux qui sont reçus à l'examen, et

4°) Il en est bien ainsi. Donc, pourrait-on conclure, il y a relativement plus de refusés en A qu'en BC parmi les refusés à l'examen. Or,

5°) C'est le contraire qui est significativement constaté !

Il est facile de se rendre compte, en lisant les lignes une par une, que le couple A.BC n'est pas le seul à présenter quelques contradictions. Si on lit le tableau colonne par colonne, un autre fait apparaît :

1ère colonne : Deux couples seulement donnent des pourcentages de réussite comparables (A.H et BC.BP)

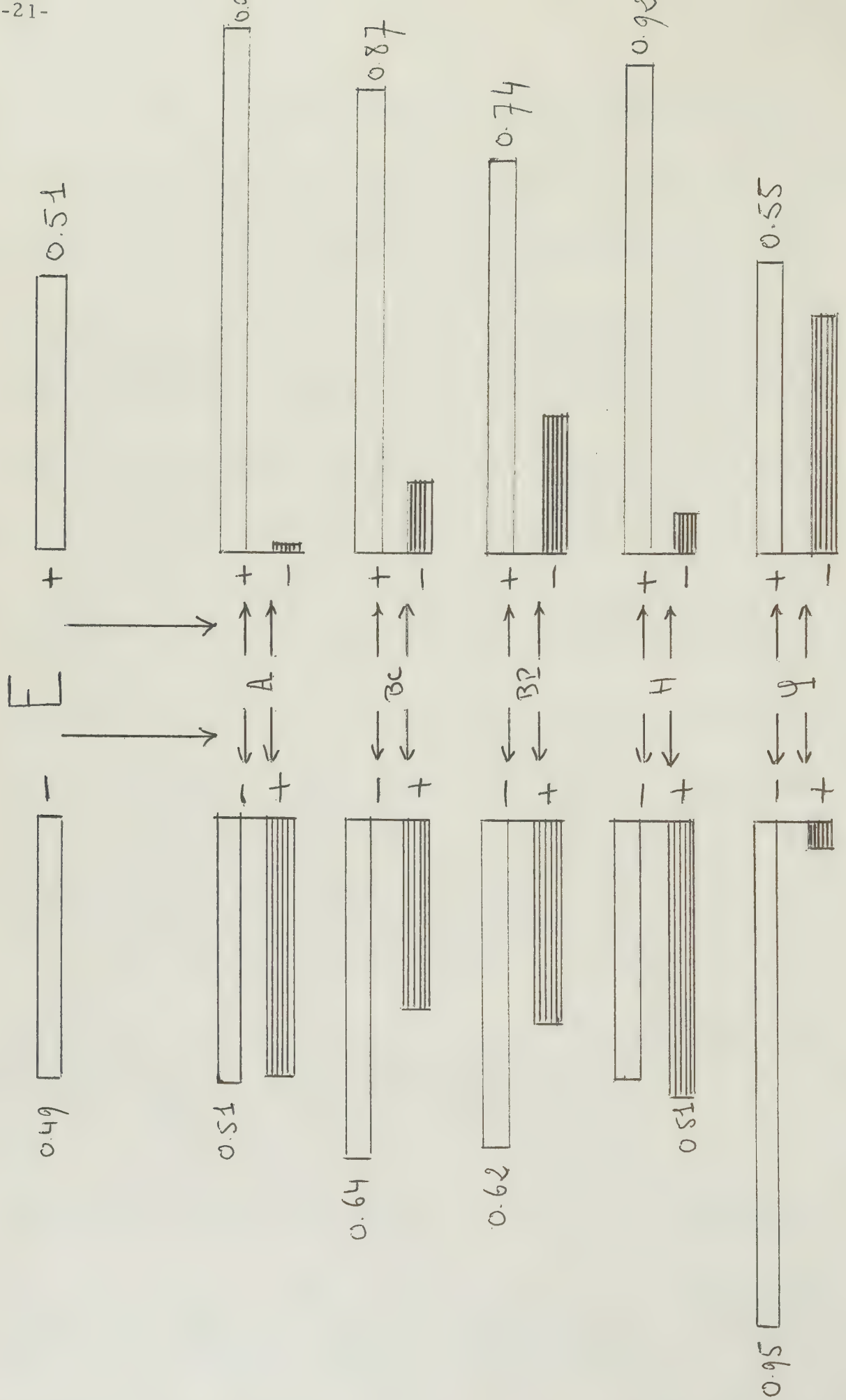
2ème colonne : Tous les couples sont mis en accord par le jugement de l'examen sauf les quatre qui comportent PHY mais, en ce qui concerne cette discipline, tous les désaccords sont négatifs (et significatifs) alors qu'au niveau de la première colonne ils étaient positifs et hautement significatifs. C'est donc qu'avant la "sanction" de l'examen on avait toujours p... > pPHY et que, après cette "sanction" pPHY est devenu supérieur. Et pour cause : 95 p.100 des reçus en PHY le sont à l'examen. PHY "place bien" ses reçus !

3ème colonne : Le plus souvent, les valeurs sont positives (huit fois sur dix et sept fois significativement). Ce qui veut dire que, les disciplines qui acceptent le plus sont finalement les plus récusées par l'examen. Et pourtant, l'examen, ce n'est rien d'autre que l'ensemble des disciplines. Les deux négatifs ne sont d'ailleurs que des confirmations du premièrement.

4ème colonne : Dix fois sur dix les écarts initiaux sont confirmés dans leur sens et cinq fois dans la grandeur de leur signification.

5ème colonne : Par rapport à la première comparaison (1ère colonne) les différences entre disciplines d'un couple sont constamment vérifiées (au signe près : ce qui est normal puisque le sens de la différence est inversé).

Par conséquent, tous les couples de disciplines présentent des caractéristiques comparables, qu'ils soient pris isolément ou en tant que contrôle de l'examen lui-même (comparaison des colonnes 1 et 4 et sur-

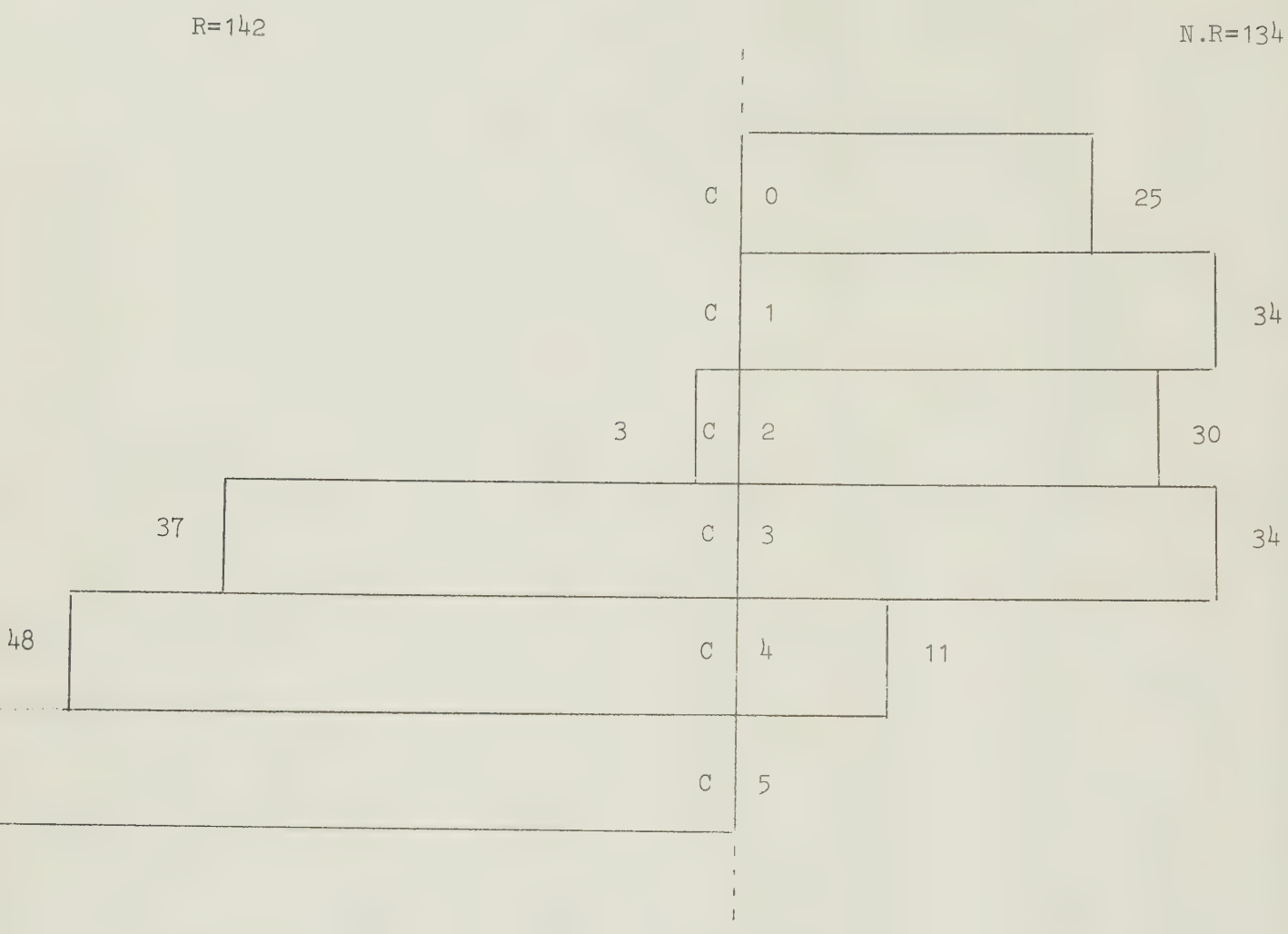


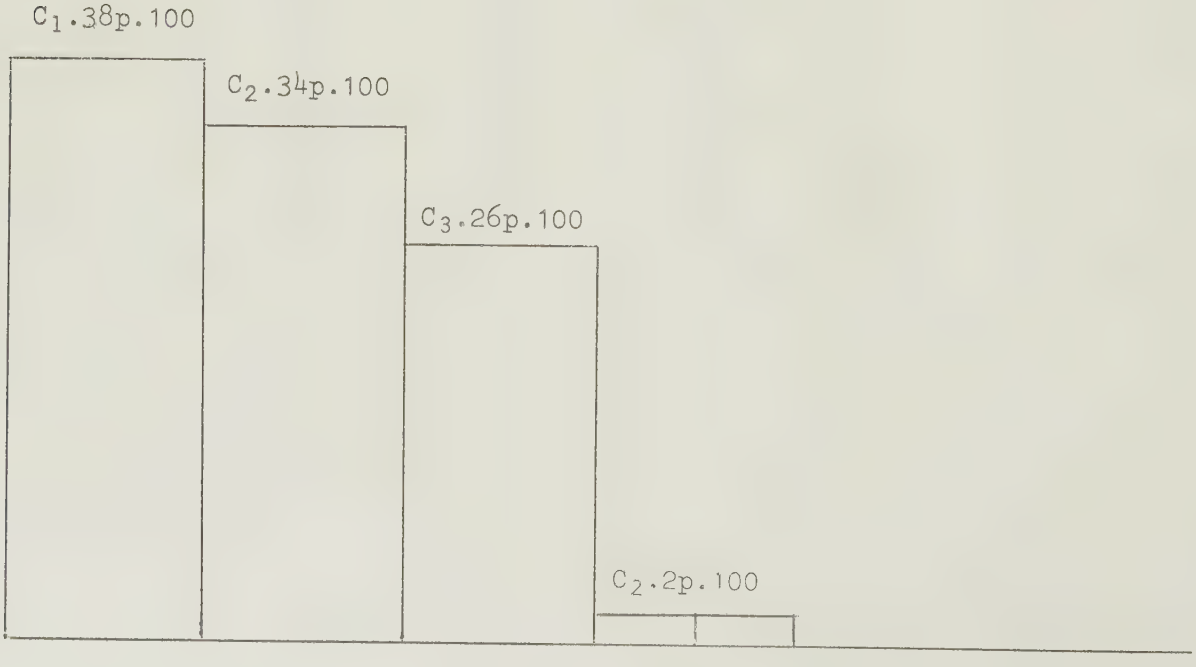
RAPHIQUE N° 8:entrée EXAMEN.

tout 1 et 5). L'examen ne sanctionne pas les différences que révèlent les couples (comparaison des colonnes 1 et 2). Enfin, l'examen modifie beaucoup plus les proportions d'échec que de réussite (comparaison des colonnes 1 et 3). Tout se passe comme si les disciplines savaient accepter mais pas refuser. Mais alors, savent-elles accepter ? L'ensemble des disciplines constituant l'examen apparaît, en gros, comme une "machine à refuser", ce que chacune d'entre elles n'a pas su faire. Les cinq disciplines formant l'examen forment un tout incohérent.

L'examen est incohérent.

Répartition des 6 classes (C_0 à C_5) en fréquences absolues.





Participation (en p.100) des différentes classes dans l'effectif des reçus.

S Y N T H E S E

Si on fait le tableau des résultats en binaire, en représentant chaque résultat par "1" s'il est favorable et "0" s'il ne l'est pas, chaque candidat peut " écrire son message" de 32 façons différentes (2^5). On peut classer ce message en sous-classes, selon le nombre de "1" qu'il présente :

. Classe 5 : 11111
. Classe 4 : 1111 et un "0" en position quelconque.
. Classe 3 : 111 et deux "0" en position quelconque.
. Classe 2 : 11 et trois"0" en position quelconque.
. Classe 1 : 1 et quatre "0" en position quelconque.
. Classe 0 : 00000

Alors que la discipline qui reçoit le moins (PHY) accepte 85 candidats, il n'y a que 54 "11111".

Alors que la discipline qui refuse le moins (A) refuse 69 candidats, il n'y a que 25 "00000".

Donc, sur 276 candidats, les cinq disciplines ne sont en accord parfait que 79 fois (environ 25 p.100) et environ deux fois plus souvent pour accepter que pour refuser.

Pour 197 candidats (environ 75 p.100) les différentes disciplines sont en désaccord. Pour 59 de ces 197 candidats (environ 30 p.100) le désaccord ne porte que sur une seule discipline : parmi ceux-ci, certains sont reçus (48), d'autres non (11). C'est naturellement sur les notes que se fonde une telle discrimination. Si l'examen est incohérent n'est-ce pas à cause des notes ?

Une dernière question se posera en ce qui concerne les effectifs : les chiffres fournis en 1970 sont-ils épisodiques ou reflètent-ils une tendance habituelle pour chaque discipline ? Voici les pourcentages des acceptés, discipline par discipline, en 69, en 70 et en 71.

E V O L U T I O N - S I T U A T I O N en 7 1

	69	70	71
A	89	75	63
BC	83	62	67
BP	87	57	65
H	89	72	72
PHY	46	31	41
Examen	81	51	62

ordre	69	70	71
A	1	1	3
BC	4	3	4
BP	3	4	2
H	1	2	1
PHY	5	5	5
examen	3>.<4	4>·<5	2 >·< 3

A l'évidence, tous les pourcentages de 69 diffèrent significativement des pourcentages des suivantes. Mais trois disciplines n'ont pas significativement changé leurs résultats entre 70 et 71 ; ce sont :

. BC : $\varepsilon = 1{,}19$ (écart réduit de la variation)
. BP : $\varepsilon = 1{,}90$ (écart réduit de la variation)
. H : $\varepsilon = 0$ (écart réduit de la variation)

La plus grande sévérité de A et la moins grande sévérité de PHY sont significatives. Cependant que le résultat final a été amélioré significativement. Ainsi est concrétisée l'approbation par l'examen des résultats de PHY et la contestation de ceux de A et, par là même, le côté relativement durable de l'incohérence de l'examen. Suffisament durable pour qu'on soit tenter d'y remédier.

On peut même démontrer que loin de s'améliorer de 70 à 71, l'incohérence s'aggrave. Les hypothèses suivantes sont fondées sur les chiffres de 71.

Première hypothèse :

La différence entre le nombre d'acceptés dans les différentes disciplines n'est pas significative. Ceci revient à dire qu'il faut chercher, pour chaque discipline, si le pourcentage de reçus par elle correspond au pourcentage de reçus au total.

a) Le pourcentage théorique des acceptés est celui de A (p)

p - pBC	=	0,06	:	ε =	1,880	Hypothèse ACCEPTEE
p - pBP	=	0,02	:	ε =	0,660	Hypothèse ACCEPTEE
p - pH	=	0,09	:	ε =	2,970	Hypothèse rejetée ($\alpha < 10^{-2}_{9}$)
p - pPHY	=	0,22	:	ε =	7,262	Hypothèse rejetée ($\alpha < 10^{-9}$)

b) Le pourcentage théorique des acceptés est celui de BC (p)

p - pA	=	0,06	:	ε =	1,931	Hypothèse ACCEPTEE
p - BP	=	0,08	:	ε =	2,575	Hypothèse rejetée ($\alpha \simeq 10^{-2}_{5}$)
p - pH	=	0,15	:	ε =	4,828	Hypothèse rejetée ($\alpha < 10^{-6}$)
p - pPHY	=	0,16	:	ε =	5,150	Hypothèse rejetée ($\alpha < 10^{-9}$)

c) Le pourcentage théorique des acceptés est celui de BP (p)

p - pA	=	0,02	:	ε =	0,668	Hypothèse ACCEPTEE
p - pBC	=	0,08	:	ε =	2,673	Hypothèse rejetée ($\alpha < 10^{-2}$)
p - pH	=	0,07	:	ε =	2,338	Hypothèse rejetée ($\alpha \simeq 0,02$)
p - pPHY	=	0,24	:	ε =	8,019	Hypothèse rejetée ($\alpha < 10^{-9}$)

d) Le pourcentage théorique des acceptés est celui de H (p)

p - pA	=	0,09	:	ε =	3,194	Hypothèse rejetée ($\alpha \simeq 10^{-3}_{7}$)
p - pBC	=	0,15	:	ε =	5,324	Hypothèse rejetée ($\alpha < 10^{-}$)
p - pBP	=	0,07	:	ε =	2,484	Hypothèse rejetée ($\alpha \simeq 0,02$)
p - pPHY	=	0,31	:	ε =	11,003	Hypothèse rejetée ($\alpha \ll 10^{-9}$)

e) Le pourcentage théorique des acceptés est celui de PHY (p)

p - pA	=	0,22	:	ε =	7,128	Hypothèse rejetée ($\alpha < 10^{-9}_{9}$)
p - pBC	=	0,16	:	ε =	5,184	Hypothèse rejetée ($\alpha < 10^{-9}_{9}$)
p - pBP	=	0,24	:	ε =	7,776	Hypothèse rejetée ($\alpha < 10^{-9}_{9}$)
p - pH	=	0,31	:	ε =	10,045	Hypothèse rejetée ($\alpha < 10^{-9}$)

BC et BP sont restées liées, A s'est liée à BC et à BP. Le couple AH a divorcé, PHY s'écarte toujours très significativement des autres.

Au total, le résultat de chaque discipline comparé à ceux de l'examen :

1°) ne diffèrent pas : A.E, BC.E, BP.E (ε respectifs : 0,330 ; 1,609 ; 1,002)

2°) diffèrent : H.E et PHY.E (ε respectifs : 3,549 et 6,804)

Deuxième hypothèse :

Les non-reçus dans une discipline sont reçus dans une autre en fonction d'un jeu de pile ou face (le pourcentage théorique est ici 0,5)

a) Lorsqu'on est refusé en A on a une chance sur deux d'être accepté ailleurs.

0,5 - pBC	=	0,16	ε	=	3,102 Hypothèse refusée ($\alpha < 10^{-3}$)
0,5 - pBP	=	0,00	ε	=	0 Hypothèse ACCEPTEE
0,5 - pH	=	0,01	ε	=	1,939 Hypothèse ACCEPTEE
0,5 - pPHY	=	0,33	ε	=	6,398 Hypothèse refusée ($\alpha < 10^{-7}$)

b) Lorsqu'on est refusé en BC on a une chance sur deux d'être accepté ailleurs.

0,5 - pA	=	0,07	ε	=	1,461 Hypothèse ACCEPTEE
0,5 - pBP	=	0,02	ε	=	0,417 Hypothèse ACCEPTEE
0,5 - pH	=	0,04	ε	=	0,835 Hypothèse ACCEPTEE
0,5 - pPHY	=	0,29	ε	=	6,055 Hypothèse refusée ($\alpha < 10^{-3}$)

c) Lorsqu'on est refusé en BP on a une chance sur deux d'être accepté ailleurs.

0,5 - pA	=	0,03	ε	=	0,566 Hypothèse ACCEPTEE
0,5 - pBC	=	0,12	ε	=	2,64 Hypothèse refusée ($\alpha < 0,03$)
0,5 - pH	=	0,03	ε	=	0,566 Hypothèse ACCEPTEE
0,5 - pPHY	=	0,24	ε	=	4,528 Hypothèse refusée ($\alpha < 10^{-5}$)

d) Lorsqu'on est refusé en H on a une chance sur deux d'être accepté ailleurs.

0,5 - pA	=	0,15	ε	=	2,939 Hypothèse ACCEPTEE
0,5 - pBC	=	0,20	ε	=	3,919 Hypothèse refusée ($\alpha < 10^{-4}$)
0,5 - pBP	=	0,09	ε	=	1,763 Hypothèse ACCEPTEE
0,5 - pPHY	=	0,37	ε	=	7,250 Hypothèse refusée ($\alpha < 10^{-9}$)

e) Lorsqu'on est refusé en PHY on a une chance sur deux d'être accepté ailleurs.

0,5 - pA	=	0,02	ε	=	0,488 Hypothèse ACCEPTEE
p - pBC	=	0,09	ε	=	2,197 Hypothèse ACCEPTEE
0,5 - pBP	=	0,06	ε	=	1,464 Hypothèse ACCEPTEE
0,5 - pH	=	0,08	ε	=	1,953 Hypothèse ACCEPTEE

Dans 13 des 20 cas examinés l'hypothèse du traitement secondaire par "pile ou face" des candidats initialement refusés ne peut être rejetée !

Si on formule la même hypothèse pour l'examen lui-même, le résultat n'est guère plus encourageant :

a) Lorsqu'on est refusé en (...) on a une chance sur deux d'être accepté à l'examen.

$pA^{-}E^{-}$	- 0,5	=	0,29	ε	=	5,623 Hypothèse refusée ($\alpha < 10^{-7}$)
$pBC^{-}E^{-}$	- 0,5	=	0,16	ε	=	3,340 hypothèse refusée ($\alpha < 10^{-3}$)
$pBP^{-}E^{-}$	- 0,5	=	0,13	ε	=	2,452 Hypothèse refusée ($\alpha < 0,02$)
$pH^{-}E^{-}$	- 0,5	=	0,27	ε	=	5,290 Hypothèse refusée ($\alpha < 0,06$)
$pPHY^{-}E^{-}$	- 0,5	=	0,12	ε	=	2,929 Hypothèse ACCEPTEE (limite)

b) Lorsqu'on est refusé en E, on a une chance sur deux d'être accepté ailleurs :

$0,5 - pA^{-}E^{-}$	=	0,27	ε	=	4,550 Hypothèse refusée ($\alpha < 10^{-5}_{4}$)
$0,5 - pBC^{-}E^{-}$	=	0,25	ε	=	4,213 Hypothèse refusée ($\alpha < 10^{-}$)
$0,5 - pBP^{-}E^{-}$	=	0,08	ε	=	1,348 Hypothèse ACCEPTEE
$0,5 - pH^{-}E^{-}$	=	0,07	ε	=	1,179 Hypothèse ACCEPTEE
$0,5 - pPHY^{-}E^{-}$	=	0,46	ε	=	7,752 Hypothèse refusée ($\alpha < 10^{-9}$)

On peut remarquer que seule PHY n'est jamais convaincue de jouer à pile ou face lorsqu'il s'agit de juger, en quelque sorte, en appel, des refusés et que cette discipline est la seule pour qui, unanimement, on puisse convaincre ses partenaires de jouer ce jeu vis-à-vis de ses propres refusés.

C O N C L U S I O N S

D'une discipline à l'autre, les effectifs déclarés satisfaisants sont différents et, d'une année à l'autre, cette incohérence ne s'améliore pas. La question qui se pose est de savoir pourquoi les résultats sont incohérents. On peut avancer l'hypothèse que chaque discipline a son propre "génie", qu'on peut être particulièrement "doué" pour telle ou telle et qu'on peut être, par nature, plutôt bon Physiologiste en même temps que mauvais Chimiste. Accepter un tel fait serait grave, car le niveau logique de l'enseignement et l'exigence normale de l'examen à un tel niveau d'étude ne doivent pas être ceux de spécialistes : ces enseignements ne sont que des initiations à des études dont aucune de ces disciplines ne représente une image fidèle. L'impression que fournit une analyse très superficielle des résultats de ces examens, est qu'ils ne sont pas destinés à prouver que les candidats répondent à des normes définies. C'est peut-être là, la cause de l'incertitude des résultats, décelée, sur ce fait surprenant que les désaccords entre les disciplines prises deux à deux sont sensiblement amoindris si, dans chacune d'elles, les effectifs sont "pondérés" par le résultat global de l'examen : une cumulation de désaccord tempérant les désaccords suggère l'idée d'une certaine inorganisation.

Enfin, il faut souligner le pourcentage extrêmement bas d'accords "globaux", de l'ordre de 25 p.100. Une fois sur quatre seulement, les cinq disciplines sont en accord soit pour accepter, soit pour refuser ; encore faut-il rappeler que les accords parfaits pour recevoir sont beaucoup plus nombreux que pour refuser (rapport du simple au double). Cet aspect des choses se retrouve au niveau de la plupart des couples de disciplines : chacune se montre relativement plus cohérente pour accepter que pour refuser, ce qui ne peut que confirmer l'impression d'incohérence de l'ensemble des opérations.

En n'oubliant certes pas que, telle ou telle discipline, d'aspect rébarbatif pour certains candidats qui, en conséquence, la négligeront d'autant plus volontiers qu'un bon résultat dans la discipline préférée peut compenser des insuffisances dans la discipline fatigante, en n'oubliant pas qu'une "impasse" organisée est d'autant plus admissible que la réglementation l'encourage, un fait reste surprenant : l'éventualité la plus fréquemment observée est celle de la double impasse (26 p.100 en 1970). Il semble difficile d'admettre que le plus grand nombre de candidats est choisi de "sacrifier" deux disciplines. Ce seul fait rend suspects les résultats et souligne leur incertitude.

Finalement, c'est peut être dans la manière de juger que quelque chose ne va pas. C'est dans cette perspective que l'expression du jugement : la note, mérite attention.

CHAPITRE II

NOTES

D'abord rumeur légère, l'absurde fiction fait plus d'une victime et l'on voit maints pauvres diables, sous cette arme redoutable, tomber terrassés.

- VALEURS 33
- COMPARAISONS 44
- DISTRIBUTION 49
- INCERTITUDE 5I
- PEUT-ON FAIRE CONFIANCE AUX NOTES ? 54
- CORRELATIONS 56

V A L E U R S

Un examen est passé. Chacune des cinq épreuves qui le composaient a été subie par 276 candidats. Les moyennes pour chacune d'entre elles ont été :

A	11,4257	11,3
BC	10,5615	10,22
BP	10,0399	9,88
H	11,0063	10,86
PHY ...	7,9057	7,92

Sauf spécifications contraires, les moyennes qui seront utilisées sont celles que fournissent les histogrammes (les secondes). Il n'y a pas d'écart significatif entre les moyennes "vraies" (les premières citées) et celles que fournissent les histogrammes. Les notes sont arrondies au point le plus proche pour n,25 ou n,75 et au point inférieur pour n,5. Enfin, la discipline qui a été réellement notée sur 30 (PHY) est ramenée à 20. Ce sont les moyennes fournies par les histogrammes qui serviront de base de travail habituelle.

L'application de deux écarts types à la moyenne observée (extraite des histogrammes) donne les écarts suivants :

	$\mu - 2S$	μ	$\mu + 2S$
A	1,22	11,3	21,38
BC	3,14	10,22	17,30
BP	3,94	9,88	15,82
H	5,58	10,86	16,14
PHY	-0,26	7,92	16,10

On peut vérifier que, compte tenu de l'arrondissement des notes, la distribution ne s'éloigne pas trop de la loi "normale".

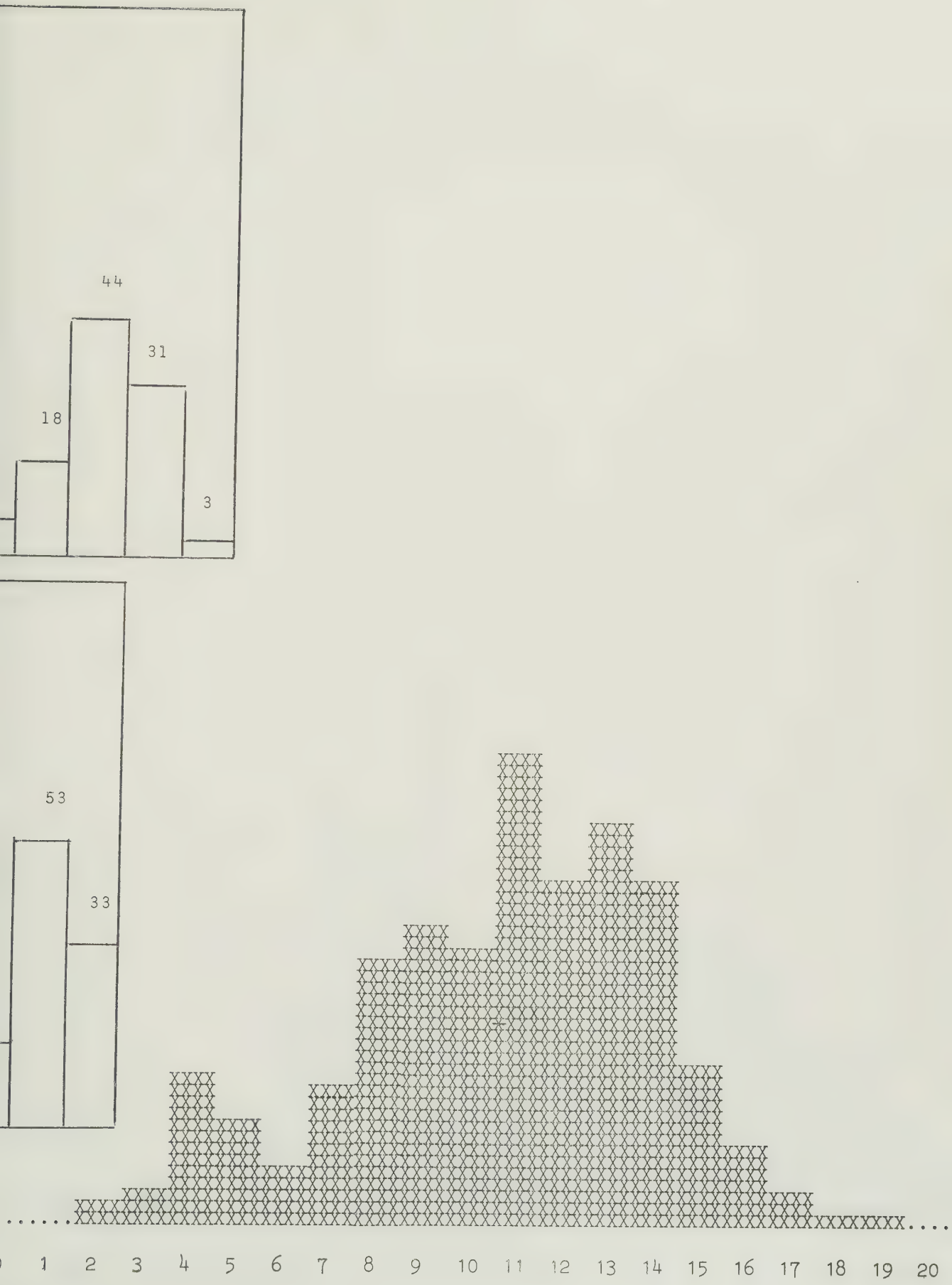
HISTOGRAMME N° I (A 6 70)

STOGRAMME N° Ibis (A 6 71)

HISTOGRAMME N° II (BC 6 70)

44
30
22
1
55
23
..XXXX

HISTOGRAMME N° III (H 6 70)

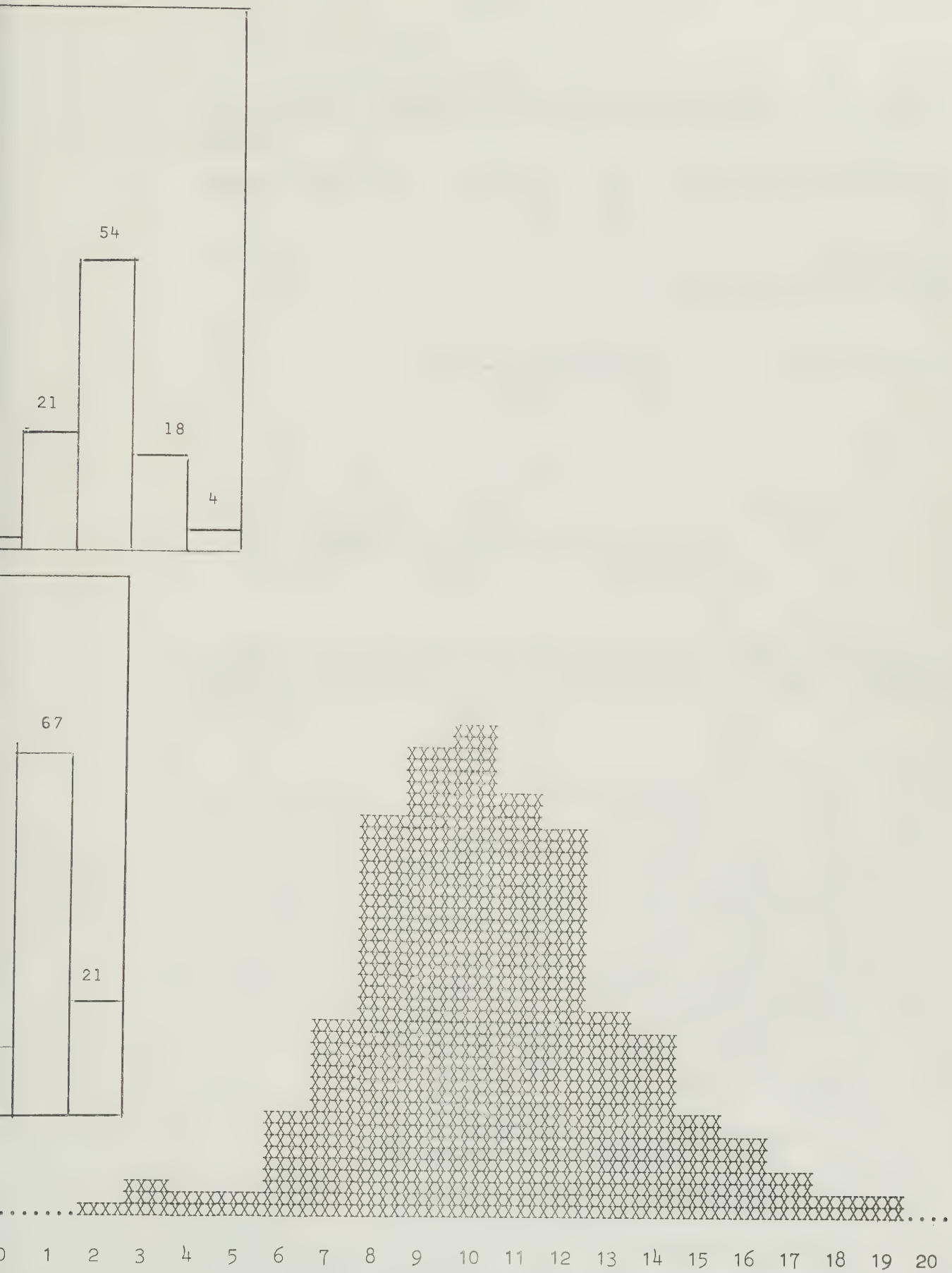


ISTOGRAMME N° III bis (H 6 71)

HISTOGRAMME N° IV (PHY. 6 70)

ISTOGRAMME N° IV bis (PHY.6 71)

ISTOGRAMME N° V (BP 6 70)



HISTOGRAMME N° V bis (BP 6 71)

C O M P A R A I S O N S

Si les histogrammes présentent quelques différences évidentes, confirmées par les valeurs des écarts types quelques questions se posent.

PREMIERE QUESTION : La différence entre les moyennes observées est-elle significative ?

Entre deux moyennes $\bar{X}$ et $\bar{Y}$, l'écart réduit sera ainsi calculé :

$$\varepsilon = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{S^2 x}{nx} + \frac{S^2 y}{ny}}}$$

où S^2 représente la variance.
nx l'effectif des candidats (ici $nx = ny$)
$\frac{S^2x}{nx}$ est la variance S'^2x de la moyenne de X.

Le tableau joint donne ces différentes valeurs. Comme l'habitude en a été prise, une valeur encadrée de ε indique que l'écart n'est pas significatif, le (-) indique que X < Y.

	ε	α
A.BC	2,91	$< 10^{-2}$
A.BP	4,03	$< 10^{-4}$
A.H	1,28	
A.PHY	8,66	$< 10^{-9}$
BC.BP	1,21	
BC.H	- 2,40	< 0,02
BC.PHY	7,07	$< 10^{-9}$
BP.H	- 4,04	$< 10^{-4}$
BP.PHY	6,44	$< 10^{-9}$
H.PHY	10,03	$< 10^{-9}$

Deux remarques :

1°) Si on calcule l'écart réduit sur les valeurs "vraies" de la moyenne, ε change peu sauf pour le couple BC.H où il prend la valeur 1,70 au lieu de 2,40. Ceci explique que sa valeur ait été encadrée bien qu'elle soit supérieure à 2.

2°) Si on compare la moyenne exacte de A (sur 20) et la moyenne exacte de PHY (sur 30) la différence n'est pas significative ! (ε = 0,93)

Dans l'ensemble les moyennes diffèrent significativement sauf celles des couples A.H, BC.BP et BC.H. Les moyennes de PHY s'écartent des autres de manière particulièrement significative.

DEUXIEME QUESTION : Dans chaque discipline, les moyennes des acceptés, des refusés et la moyenne générale diffèrent-elles ?

On désignera par :
C^+x les candidats acceptés dans la discipline X
C^-x ceux qui ont été refusés.
La moyenne des reçus dans la discipline X se représentera donc :C^+x, celle des refusés : C^-x, la moyenne générale se représentant par $\bar{X}$.
Voici les résultats :

$\mu C^+A, \bar{A}$	: ε =	6
$\mu C^-A, \bar{A}$	: ε =	13
$\mu C^+A, \mu C^-A$	: ε =	20
$\mu C^+BC, \overline{BC}$	: ε =	8,88
$\mu C^-BC, \overline{BC}$	: ε =	11
$\mu C^+BC, \mu C^-BC$	: ε =	22
$\mu C^+BP, \overline{BP}$	: ε =	8
$\mu C^-BP, \overline{BP}$	: ε =	6
$\mu C^+BP, \mu C^-BP$	: ε =	20
$\mu C^+H, \bar{H}$	: ε =	1,86
$\mu C^-H, \bar{H}$	: ε =	14
$\mu C^+H, \mu C^-H$	: ε =	21
$\mu C^+PHY, \overline{PHY}$	: ε =	13
$\mu C^-PHY, \overline{PHY}$	: ε =	7
$\mu C^+PHY, \mu C^-PHY$	: ε =	20

Remarques :

1°) Trois disciplines discriminent plus les refusés que les acceptés par rapport à la moyenne générale : A, BC, H

2°) Les deux autres procèdent inversement : BP et PHY

3°) Une discipline (H) qui reçoit 72 p.100 de son effectif (2ème rang)

TABLEAU 1.

	μX^+	μY^+	μX^-	μY^-	$\mu \bar{X}$	$\mu \bar{Y}$
	+		-		T	
A BC	+ 4 74		- 4 52		+ 2 49	
A BP	+ 6 76		- 6 48		+ 4 18	
A H	+ 6 54		- 7 27		+ 1 31	
A PHY	+ 2 92		- 2 64		+ 7 61	
BC BP	+ 2 68		- 2 50		+ 1 90	
BC H	+ 2 02		- 3 75		- 1 70	
BC PHY	- 0 18		+ 3 20		+ 6 26	
BP H	- 0 90		- 1 59		- 4 04	
BP PHY	- 1 90		+ 6 90		+ 5 19	
H PHY	- 1 39		+ 8 29		+ 7 71	

TROISIEME QUESTION : Dans chaque catégorie "acceptés","refusés" et "total" les disciplines présentent-elles des moyennes significativement différentes ? (cf tableau 1)

1°) En ce qui concerne les acceptés :

Oui pour A dans tous ses rapports. Mais, et c'est étonnant, on trouve que PHY, jusqu'ici jamais d'accord, le devient avec presque tout le monde, et, même en sollicitant un peu les chiffres et compte tenu de ce qui a précédé, avec A. Enfin et surtout, H et PHY se rencontrent alors que l'écart de leur pourcentage de reçus était significatif avec un risque très inférieur à 1/1000000000 !

2°) Pour les refusés :

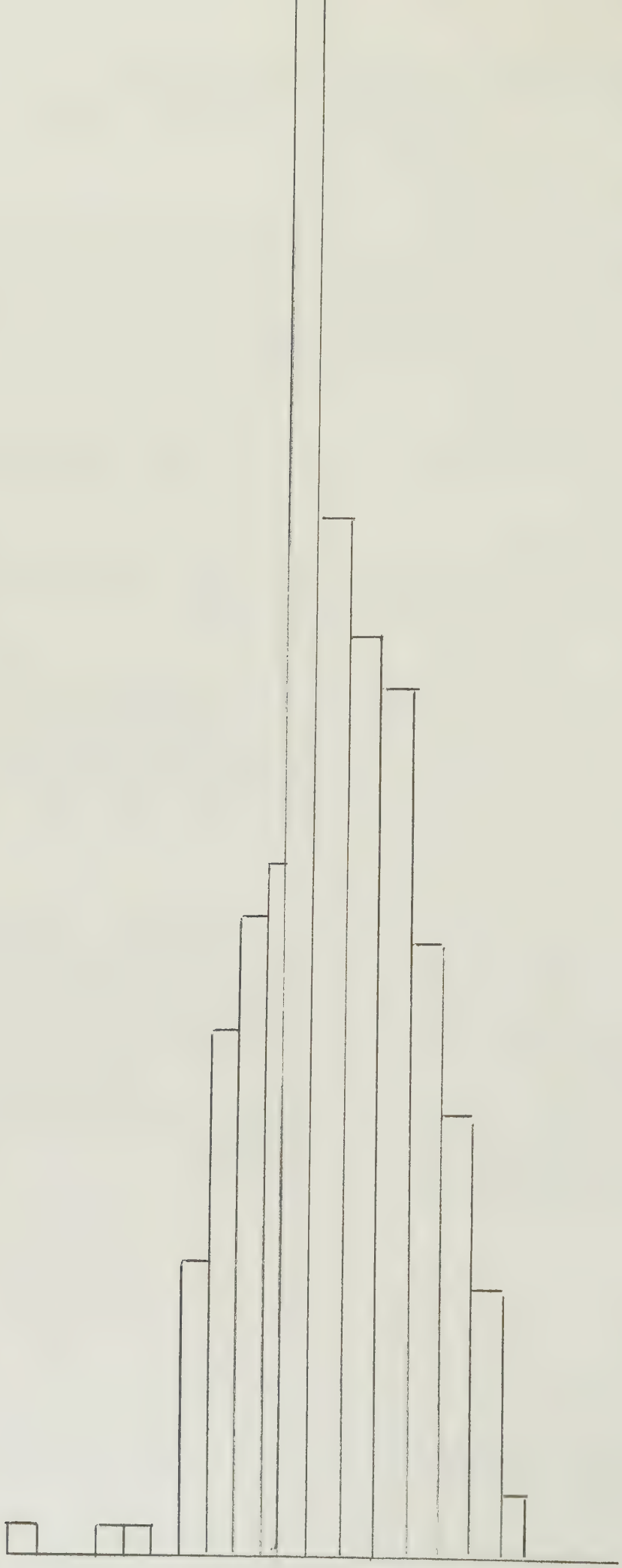
Les six premiers couples (de A.BC à BC.H) sont satisfaisants pour l'esprit : le signe des ε est le contraire pour les refusés, de celui des reçus.

On voit que la moyenne des refusés en BC s'éloigne significativement de celle de PHY, que BP et H ne diffèrent pas significativement, ce qui n'a rien d'étonnant lorsqu'on jette ne serait-ce qu'un coup d'oeil rapide sur les histogrammes correspondants.

PHY a significativement une moyenne inférieure à celles de BC, BP et H (puisque ces ε sont "positifs") mais on peut relever avec étonnement que la moyenne des refusés en PHY est supérieure à celle de A. A cela une explication simple : la quantité (69 p.100 de l'effectif de PHY est situé au-dessous de 10, et donc refusé, mais tous n'ont pas un ou deux, alors que la plus grande partie des refusés en A est concentrée dans la partie gauche de l'histogramme).

3°) Au total,

Un couple semble exemplaire : A.H (et, à un moindre degré BC.BP). Voici un couple de deux disciplines ayant accepté un nombre presque égal de candidats : la différence de leurs pourcentages n'est que de 0,03, l'écart réduit de leur différence n'est que de 0,81 ; que l'écart réduit de leurs moyennes ne soit pas significatif, rien là qui choque la logique. Mais les écarts de leurs moyennes sont très significatifs tant pour les acceptés ($\varepsilon = 6{,}54$; $\alpha < 10^{-9}$) que pour les refusés : ($\varepsilon = 7{,}27$; $\alpha < 10^{-9}$) : c'est plus, semble-t-il, la distribution des notes qui mérite analyse que les moyennes elles-mêmes qui, bien que grandeurs commodes et faciles à calculer, apparaissent comme relativement sans signification. Or la distribution des notes, c'est la conséquence de la correction.



UN HISTOGRAMME BIMODAL !

D I S T R I B U T I O N

Il a été dit plus haut que la distribution ne s'éloignait "pas trop" de la loi normale. En effet, en dehors de l'intervalle μ - deux écarts types-μ + 2 écarts types, on doit trouver 5 p.100 de la population, c'est-à-dire, ici, environ 14 (13,8 exactement). Pour H, par exemple, on trouve en réalité 4, et c'est un peu loin de 14. Mais si on calcule l'écart type à la moyenne des "acceptés" et celui qui correspond à celle des refusés, si on compte le nombre d'individus se trouvant au-dessous de la "petite moyenne" diminuée de deux écarts-types et au-dessus de la "grande moyenne" augmentée de deux écarts-types on en dénombre 13. L'histogramme de H serait donc en réalité bi modal mais, si on le regarde, on n'en croit pas ses yeux. A moins qu'on imagine que cette impressionnante colonne de "10" ne soit qu'un artefact : certains "valant" 9 (ou moins), d'autres 11 (ou plus) et quelques uns ..... 10.

Voici les moyennes, rappelées des "acceptés" dans les différentes disciplines, leurs écarts-types et les écarts μ ± 2S.

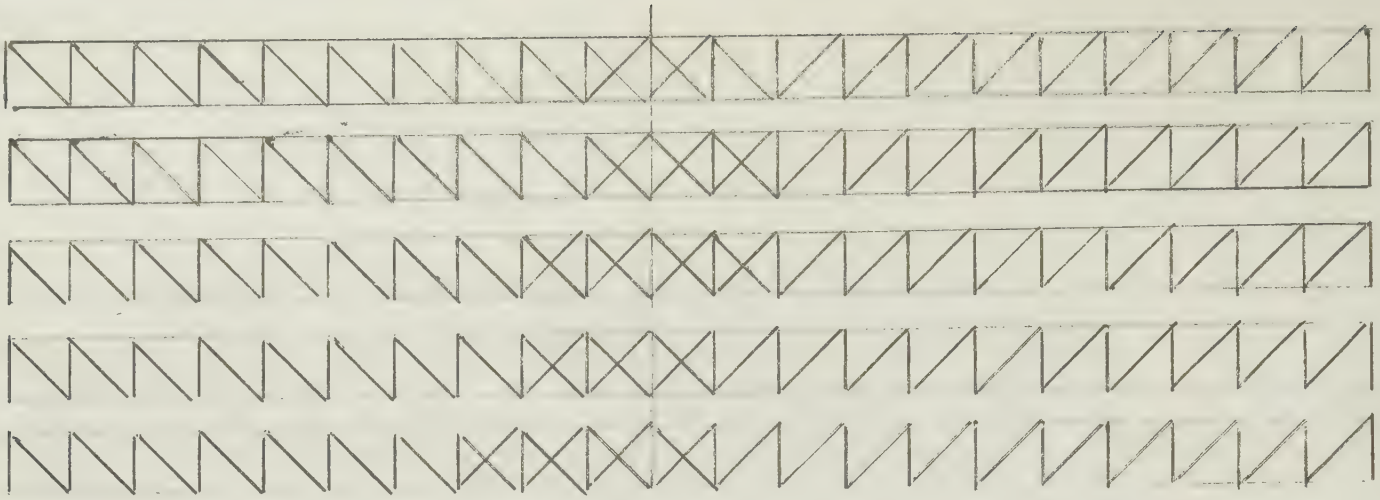
A^+	: 13,5 ± (2 x 2,48)	:	écart	de	8,54	à	18,48
BC^+	: 12,46 ± (2 x 1,76)	:	écart	de	8,94	à	15,98
BP^+	: 11,89 ± (2 x 2,05)	:	écart	de	7,79	à	15,99
H^+	: 12,08 ± (2 x 1,87)	:	écart	de	8,34	à	15,82
PHY^+	: 12,52 ± (2 x 5,48)	:	écart	de	7,04	à	18,00

Si on fait la même opération sur les moyennes des refusés ($\bar{X}^-$) on obtient :

A^-	: 4,5 ± (2 x 3,33)	:	écart	de	-2,16	à	11,16
BC^-	: 6,59 ± (2 x 2,35)	:	écart	de	1,89	à	11,29
BP^-	: 7,29 ± (2 x 1,71)	:	écart	de	0,90	à	10,71
H^-	: 7,65 ± (2 x 1,42)	:	écart	de	2,17	à	10,49
PHY^-	: 5,66 ± (2 x 2,40)	:	écart	de	0,86	à	10,46

On peut déjà remarquer que les notes représentant les extrémités des écarts "se tiennent" et aussi que, pour chaque discipline, elles se "recouvrent". Si on rapproche, pour chaque discipline, la note maximum que peut espérer un refusé ($\mu\bar{X}^- + 2S^-$) et celle que risque raisonnablement, au minimum, un accepté ($\mu\bar{X}^+ - 2S^+$) voici ce qu'on obtient :

A^-	max. :	4,5 + 6,66 =	11,16 soit environ		11
A^+	min. :	13,5 - 4,96 =	8,54 soit environ		9
BC^-	max. :	6,59 + 4,70 =	11,29 soit environ		11
BC^+	min. :	12,46 - 3,52 =	8,94 soit environ		9
BP^-	max. :	7,29 + 3,42 =	10,71 soit environ		11
BP^+	min. :	11,89 - 4,1 =	7,79 soit environ		8
H^-	max. :	7,65 + 2,84 =	10,49 soit environ		10
H^+	min. :	12,08 - 3,74 =	8,34 soit environ		8
PHY^-	max. :	5,66 + 4,80 =	10,46 soit environ		10
PHY^+	min. :	12,52 - 5,48 =	7,04 soit environ		7



Cette répartition donne les résultats suivants:

A = 207

BC= 171

BP= 156

H= 199

PHY.= 85

EXAMEN = 142

A = 156

BC = 140

BP = 156

H = 145

PHY = 142

L'ajustement des moyennes (cf ci-dessous) donnerait les résultats ci-dessus.

1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20

0 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20

0 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20

0 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20

0 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19

UN DECALAGE DES " MOYENNES " QUI HARMONISERAIT LES RESULTATS.

I N C E R T I T U D E

On pourrait, à partir de ces données, tenir ce raisonnement : la zône de recouvrement n'est pas vraiment significative, elle est incertaine. Sont certainement reçus dans la discipline, ceux qui ont une note supérieure à $\mu\bar{X}^- + 2S^-$ et certainement refusés dans la discipline, ceux qui ont une note inférieure à $\mu\bar{X}^+ - 2S^+$. Les autres sont incertains. Ce qui donnerait les zônes d'incertitude suivantes :

	notes incertaines	refusés à	reçus à	effectifs incertains
A	9 11	8 et moins	12 et plus	54
BC	9 11	8 et moins	12 et plus	90
BP	8 11	7 et moins	12 et plus	154
H	8 10	7 et moins	11 et plus	101
PHY	7 10	6 et moins	11 et plus	96

Tout se passe d'ailleurs comme si cette règle avait été appliquée. L'examen a reçu 142 candidats, c'est-à-dire entre 125 et 159 ($\bar{n} \pm 2S$). Avec quelle note chaque discipline peut-elle "placer" ses candidats dans cette fourchette ?

En A, note au moins égale à 12 : effectif correspondant 156
note au moins égale à 13 : effectif correspondant 126

En BC , note au moins égale à 11 : effectif correspondant 140

En BP , note au moins égale à 10 : effectif correspondant 156

En H , note au moins égale à 11 : effectif correspondant 145

En PHY, note au moins égale à 8 : effectif correspondant 142

Le choix de A	s'est fait AU-DESSUS de la zône d'incertitude	9.... 11
Le choix de BC	s'est fait à la limite supérieure de la zône d'incertitude	9 ... 11
Le choix de BP	s'est fait en DEDANS de la zône d'incertitude	8 ... 11
Le choix de H	s'est fait AU-DESSUS de la zône d'incertitude	8 ... 10
Le choix de PHY	s'est fait en DEDANS de la zône d'incertitude	7 ... 10

Mais le fait de choisir des notes "de réception" cohérentes n'impliquerait pas nécessairement que le nombre de reçus au total (c'est à dire à l'examen) soit compris dans la fourchette prévue. C'est alors les candidats qui seraient incohérents et leur élimination serait plus justifiée.

En conclusion, on pourrait représenter, pour chaque discipline, un HISTOGRAMME REDUIT faisant apparaître trois catégories de candidats : les refusés, les reçus et les INCERTAINS.

Voici ce que cela donnerait (en fréquence absolue)

	Refusés	Incertains	Reçus
A	60	54	162
BC	79	90	117
BP	51	154	71
H	30	101	145
PHY	107	96	73

Il est remarquable que le nombre d'incertains est parfois le plus important (BP).

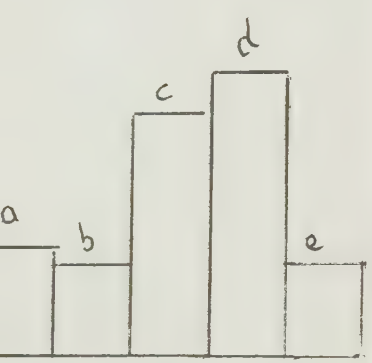
a = 0-1-2-3-4

b = 5-6-7-8

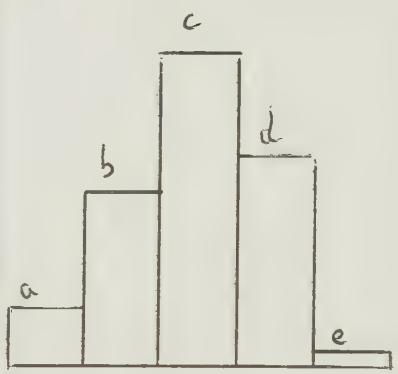
c = 9 10-11-12

d = 13-14-15-16

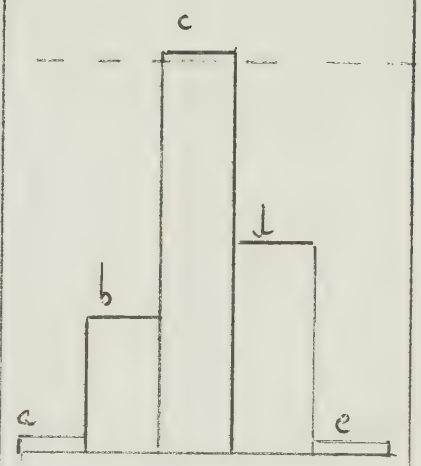
e = 17-18-19-20-



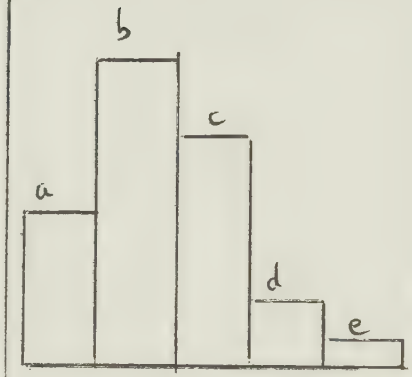
A



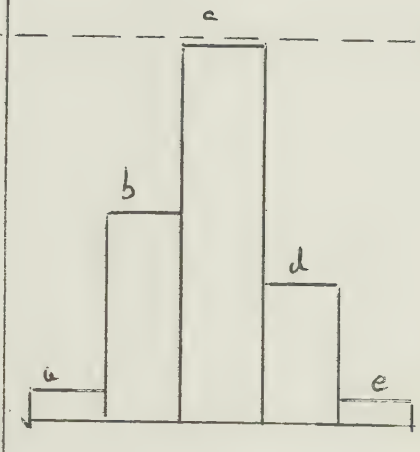
BC



H



Ψ

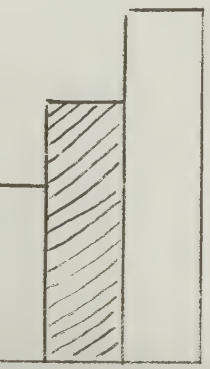


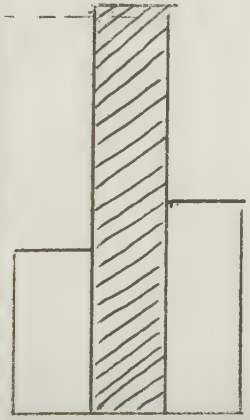
BP

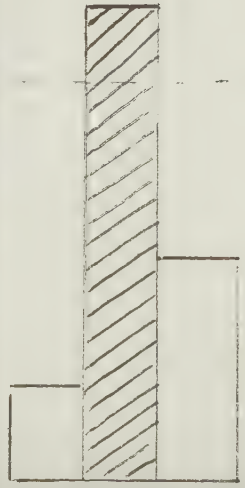
50p100

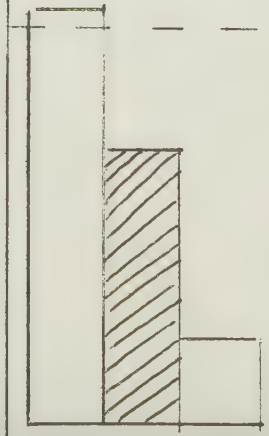
Hachures = notes donnant des résultats INCERTAINS

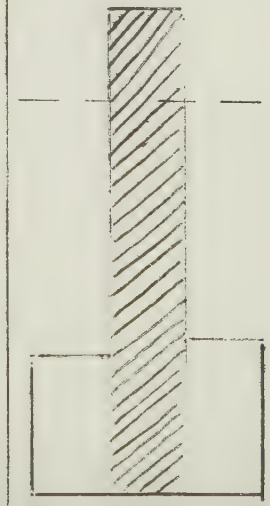
[8-12]











50p100

PEUT-ON FAIRE CONFIANCE AUX NOTES ?

Dans une large mesure, non. La prétendue "moyenne" qui est, en réalité, la médiane des possibilités de distribution offerte aux correcteurs, fausse tout. D'abord, il est évident que si une erreur de 10 p.100 est accordée sur la notation c'est précisément au niveau de cette "moyenne" (en réalité médiane) qu'elle sera plus importante. 10 p.100 de 2 n'est pas considérable. 10 p.100 de 10 cela donne 9 ou 10, ou 11. A 9 on est refusé, à 11 on est accepté. Pour deux disciplines, 10 est la note la plus fréquemment attribuée, et c'est justement la plus incertaine. Une erreur de même envergure appliquée à 18 sur 20 a beaucoup moins d'importance : 18 ± 1,8 cela fait toujours une note largement suffisante pour être reçu. L'incertitude n'est pas imaginaire : Il était indiqué que 8 en PHY valait 12 en A et peut-être 13. Si on applique systématiquement la même règle aux cinq disciplines, on regroupe les valeurs ainsi :

. si on appelle *a* les notes : 0, 1, 2, 3, 4 sur 20
b les notes : 5, 6, 7, 8, sur 20
c les notes : 9, 10, 11, 12 sur 20
d les notes : 13, 14, 15, 16 sur 20
e les notes : 17, 18, 19, 20 sur 20

Le regroupement de ces notes figure les "histogrammes réduits" exprimés en fréquences relatives. Pour simplifier, on peut regrouper $(a + b)$ et $(d + e)$: les "certainement refusés" et les "certainement acceptés". Reste c : les incertains. Et dans trois disciplines sur cinq ce sont les plus nombreux.

L'expression finale de l'examen est une somme de notes. Les notes incertaines sont les plus nombreuses (environ 49 p.100 de l'ensemble des notes). Donc la note finale est, pour une large part, une SOMME D'INCERTITUDE. Sur elle, le sort des candidats est scellé.

	ε de pX-pY	risque corresp.	rang du risque des ε	rang de corrélat. (r)	valeur de r	valeur de t
A BC	3 33	$< 10^{-3}$	4	3	0 601	12 447
A BP	4 61	$< 10^{-5}$	6	8	0 385	6 922
A H	0 81		1	1	0 667	14 818
A PH	11 50	$< 10^{-9}$	10	4	0 571	11 513
BC BP	1 21		2	10	0 284	4 902
BC H	2 56	< 0.02	3	2	0 645	13 971
BC PH	7 75	$< 10^{-9}$	8	6	0 461	8 599
BP H	3 75	$< 10^{-3}$	5	9	0 380	6 80
BP PH	6 50	$< 10^{-9}$	7	7	0 437	8 042
H PH	10 78	$< 10^{-9}$	9	5	0 503	9 633

TABLEAU 3

SYNTHESE

RANGS:	1	2	3	4	5	6	7	8	9	10
A	+ 0		0	+ 0		+			0	+
BC		+ 0	+ 0	+		0		+		0
BP		+			+	+	+ 0	0	0	0
H	+ 0	0	+		+ 0			0	+	
PH				0	0	0	+ 0	+	+	+

C O R R E L A T I O N S

Le tableau n° 3 rappelle la valeur de ε (écart réduit) des différences de pourcentage observé dans le nombre des "acceptés" dans chacun des dix couples possibles. La valeur de l'écart réduit est suivi de la valeur du risque (α) permettant éventuellement de rejeter l'hypothèse d'écarts non significatifs entre les pourcentages observés. Une troisième colonne représentera le "rang" du couple étudié en ce qui concerne l'écart réduit, suivie de la colonne représentant le rang du même couple en fonction de la valeur du coefficient de corrélation observé "r". Ce coefficient lui-même sera placé en cinquième colonne tandis que sa signification, évaluée par le test "t" sera indiquée en sixième colonne.

Le passage de r à t se fera en appliquant la formule :

$$t = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n - 2}$$

avec $(n - 2)$ constant : $276 - 2 = 274$

Deux remarques s'imposent d'emblée :

1°) Le couple AH, qui présentait le plus faible écart entre les pourcentages d'acceptés, présente aussi la corrélation la plus élevée entre les notes. Par ailleurs, les éléments de ce couple présentaient encore une différence non significative de leurs moyennes. Ce qui ne les empêchait pas :

a) de présenter les histogrammes les plus opposés : A note entre 0 et 20, H entre 1 et 19, mais la répartition des notes est tout à fait différente ;
b) de noter significativement de 1 à 21 pour A et de 6 à 16 pour H ;
c) d'être en désaccord complet sur la moyenne des acceptés et des refusés : l'écart réduit des acceptés est de 6,54, celui des refusés de 7,27, tous deux significatifs à un risque de 10^{-9}

2°) Le couple BC.BP, qui présentait un écart non significatif de leurs pourcentages d'acceptés, un écart non significatif de leurs moyennes, un désaccord réel, certes, mais faible entre les moyennes de leurs acceptés et de leurs refusés respectifs, présentent aussi la corrélation la moins forte entre leurs notes.

En général, toutes les corrélations sont positives et significatives. Donc, tout le monde reconnaît, en gros, les meilleurs des plus mauvais de la même façon. Huit sur dix des pourcentages d'acceptés diffèrent significativement. Or, ce qui permet de discerner un accepté d'un refusé, c'est la note : inférieure ou non à 10 sur 20. Accord sur les notes, désaccord sur les résultats, que signifie donc la note ?

Le tableau de synthèse indique le rang du couple en ce qui concerne et la valeur de l'écart réduit (ε de pX - pY) et celui qui concerne la signification de r. Le rang déterminé par la valeur de ε est représenté par (+) et celui qui est la conséquence de la valeur de r par (o).

CHAPITRE III

DES NOTES ET DES EFFECTIFS

Il est plus honnête et finalement plus efficace de considérer les choses telles qu'elles sont et non telles qu'on voudrait qu'elles soient.

Devant 71 p. 100 de désaccords entre les cinq épreuves constituant l'examen, cinq questions se posent :

1°- Chaque discipline distingue-t-elle les"certains des incertains" par une moyenne différente ? 60

2°- La certitude s'exprime-t-elle identiquement dans chaque discipline ? 62

3°- Les acceptés sont-ils jugés meilleurs et les refusés moins bons ? 64

4°- Les certainement acceptés et les certainement refusés sont-ils définis ? 66

5°- Y-a-t-il une différence entre la population des unanimement acceptés, d'une part, des unaninement refusés, d'autre part, et celle des incertains ? 67

TABLEAU BINAIRE

	1970				1971			
CLASSES	CATEGORIES	R	NR	T	CATEGORIES	R	NR	T
Cl.N°5	IIIII	54	..	54	IIIII	65	..	65
Cl.N04	IIIIO	9	1	10	IIIIO	7	..	7
	IIIOI	33	10	43	IIIOI	21	1	22
	IIOII	1	..	1	IIOII	2	..	2
	IOIII	5	....	5	IOIII	5	..	5
	OIIII	..	..	XX	OIIII	3	..	3
		48	11	59		38	1	39
Cl.N°3	IIIOO	18	14	32	IIIOO	4	3	7
	IIOIO	4	..	4	IIOIO	2	.2	2
	IOIIO	3	..	3	IOIIO	6	..	6
	IIOOI	3	2	5	IIOOI	4	1	5
	OIIIO	..	..	XX	OIIIO	2	..	2
	IOIOI	7	8	15	IOIOI	12	2	14
	IOOII	1	1	2	IOOII	1	..	1
	OIIOI	1	16	7	OIIOI	6	7	13
	OIOII	..	2	2	OIOII	1	..	1
	OOIII	..	1	1	OOIII	4	..	4
		37	34	71		42	13	55
Cl.N°2	IIOOO	1	5	6	IIOOO	3	1	4
	IOIOO	1	11	12	IOIOO	4	5	9
	OIIOO	..	2	2	OIIOO	1	4	5
	IOOIO	1	1	2	IOOIO	1	..	1
	OIOIO	..	..	XX	OIOIO	..	..	XX
	OOIIO	..	..	XX	OOIIO	1	2	3
	IOOOI	..	3	3	IOOOI	..	4	4
	OIOOI	..	3	3	OIOOI	..	2	2
	OOIOI	..	5	5	OOIOI	1	9	10
	OOOII	..	..	XX	OOOII	..	1	1
		3	30	33		11	28	39
Cl.N°1	IOOOO	..	10	10	IOOOO	1	5	6
	OIOOO	..	3	3	OIOOO	..	5	5
	OOIOO	..	10	10	OOIOO	..	8	8
	OOOIO	..	1	1	OOOIO	1	1	2
	OOOOI	..	10	10	OOOOI	..	13	13
		XX	34	34		2	32	34
Cl.N°0	OOOOO	..	25	25	OOOOO	..	22	22
	TOTAUX	142	134	276	TOTAUX	158	96	254

Un examen est passé, 276 candidats l'ont subi, chacune des cinq disciplines a fait son bilan. Dans chaque discipline certains candidats sont acceptés et d'autres refusés. La différence, à ce point de vue, d'une discipline à l'autre, n'est pas négligeable. En ignorant les notes, mais sans oublier que le caractère "accepté" est défini par une note au moins égale à 10 sur 20 (avec 9,5 = 9 et 9,51 = 10), il est possible de dresser le tableau binaire (tableau n° 1). L'ordre des colonnes est ici : A, BC, H, PHY, BP. Le caractère "accepté" est représenté par "1". Le caractère "refusé" par "0". Ainsi 32 catégories sont possibles. Si on s'intéresse à la colonne "T", qui représente l'effectif de chaque catégorie sans se préoccuper de savoir qui, finalement, est reçu ou non (R ou NR), une remarque s'impose : l'impressionnante faiblesse des concordances.

Tout le monde est d'accord pour accepter 54 candidats qui seront donc considérés comme indiscutablement reçus. L'accord est parfait pour refuser 25 candidats, considérés, par conséquent, comme indiscutablement refusés ou "non reçus". Le terme "non reçus" est préféré au terme "éliminés" d'une part parce que, administrativement, éliminé se rapporte aux étudiants (éliminé des études) et non aux candidats et, d'autre part parce que, comme il a été vu et qu'il sera répété, la non réception d'un candidat ne semble pas l'aboutissement d'un processus logique défini. Si on compte le nombre des accords dans un sens ou dans l'autre, et qu'on le retranche de 276, on trouve : 197. Les disciplines sont en désaccord 197 fois sur 276, soit 71 p.100 des cas ! L'accord est l'exception, le désaccord la règle.

Encore est-il utile de se poser un certain nombre de questions concernant les accords afin de se faire une opinion et de tenter d'émettre un jugement. Cinq questions seront envisagées, chacune posée sur les "reçus certains" (54) et sur les "non reçus certains" (25). Ce qui n'empêchera pas, ultérieurement, de tenter un jugement sur cette certitude, simplement posée ici en postulat. C'est-à-dire que la légitimité de cette certitude même sera mise en question.

PREMIERE QUESTION

Dans chaque discipline, la moyenne des "certains" est-elle significativement différente de celle des "incertains" ?

a) Qui sont les"certains" ?

Ceux qui ont été acceptés par toutes les disciplines. Ils constituent la classe "5" (5 réussites) représentée ainsi : ${}_5X$, ce qui se lit : candidat de la discipline X, placé en classe "5". Ce sont également les refusés par toutes les disciplines, classe 0, ainsi représentés ${}_0X$.

b) Les "incertains" sont donc les autres, ainsi représentés : ${}_1^4X$ ce qui se lit : candidat de la discipline X, placé dans les classes de 1 à 4.

c) Quelle réponse doit-on attendre à cette question ?

"Oui", évidemment, car cela signifie que, en tout, les candidats acceptés partout sont les meilleurs et les refusés partout les moins bons.

d) Comment répondre à cette question ?

En établissant le rapport de la différence des moyennes envisagées à la racine carrée de la somme de leurs variances. Soit :

$$\varepsilon = \frac{{}_5\bar{X} - {}_1^4\bar{X}}{\sqrt{S'^2\,{}_5X + S'^2\,{}_1^4X}} \qquad \text{ou} \qquad \varepsilon = \frac{{}_1^4\bar{X} - {}_0\bar{X}}{\sqrt{S'^2\,{}_1^4X + S'^2\,{}_0X}}$$

Chacune de ces différences doit-être positive. Si ε est inférieur à 2, la différence sera considérée comme non significative.

e) Réponses :

	${}_5C$		${}_0C$	
	ε	α	ε	α
A	10,74	10^{-9}	14,77	$< 10^{-9}$
BC	8,30	10^{-9}	10,85	$< 10^{-9}$
BP	8,31	10^{-9}	5,85	$< 10^{-9}$
H	6,42	10^{-9}	7,50	$< 10^{-9}$
PHY	13,02	10^{-9}	5,00	$< 10^{-9}$

f) Conclusion :

Toutes les différences étant positives on peut dire que, en se fiant aux moyennes :

1°) la classe 5C comporte les meilleurs,
2°) la classe oC comporte les moins bons.

La réponse fournie est la réponse attendue. L'examen est "bon" puisqu'il considère comme significativement meilleurs, les meilleurs, et significativement moins bons, les moins bons. MAIS cet accord n'est significatif que pour les effectifs concernés, soit pour la classe 5C : 19,548 p.100 (20 %) et pour la classe oC : 9,05 p.100 (9 %), soit en tout moins de 30 p. 100 !

D E U X I E M E Q U ' E S T I O N

A l'intérieur d'une même classe, la moyenne diffère-t-elle significativement d'une discipline à l'autre ?

a) Quelle réponse doit-on attendre à cette question ?
"Non". Cela signifierait que toutes les disciplines ont la même manière d'exprimer leurs jugements concernant les meilleurs et les moins bons.

b) Comment répondre à la question ?
En comparant ainsi les dix couples possibles :

$$\varepsilon = \frac{{}_5\bar{X} - {}_5\bar{Y}}{\sqrt{S'^2_{5X} + S'^2_{5X}}} \qquad \text{ou} \quad \varepsilon = \frac{{}_o\bar{X} - {}_o\bar{Y}}{\sqrt{S'^2_{oX} + S'^2_{oY}}}$$

c) Réponses :

NON = non significatif

	5C		oC	
	ε	α	ε	α
A.BC	5,50	$< 10^{-7}$	- 4,52	$< 10^{-5}$
A.BP	6,05	$< 10^{-}$	- 4,95	$< 10^{-6}$
A.H	5,06	$< 10^{-}$	- 7,18	$< 10^{-7}$
A. PHY	4,08	$< 10^{-}$	0,07 NON	
BC.BP	1,14 NON		- 1,2 NON	
BC.H	- 0,57 NON		- 2,93	$< 10^{-2}$
BC.PHY	- 0,09 NON		4,21	$< 10^{-4}$
BP.H	- 1,68 NON		- 0,98 NON	
BP.PHY	- 0,97 NON		4,70	$< 10^{-5}$
H.PHY	0,33 NON		6,59	$< 10^{-9}$

d) Conclusions :

Onze fois sur vingt, la réponse n'est pas celle qu'on attend. Quatre disciplines seulement expriment un jugement comparable en 5C : A s'en écarte notablement. A, par rapport aux autres disciplines, surestime la valeur des meilleurs. Ceci est à rapprocher du fait que la note des acceptés en A déviait de la moyenne.de + 2S de telle sorte, que le maximum extrapolé au risque 5 p.100 était supérieur à 20/20 (mais inférieur à 20/20 dans toutes les autres disciplines).

En ce qui concerne oC, les discordances sont majoritaires : l'évaluation des moins bons est différente d'une discipline à l'autre, les couples A.PHY, BC.BP, BP.H exceptés. Ce fait n'est pas une nouveauté, mais le trouver encore évident dans le minuscule groupe (9 p.100) des refusés unanimement, lui donne toute son importance.

T R O I S I E M E Q U E S T I O N

Dans une même discipline, la moyenne des acceptés ($\bar{X}^+$) comparée à celle des reçus certains ($_5\bar{X}$) ou des refusés et des non reçus ($\bar{X}^-$ et $_o\bar{X}$) diffèrent-elles significativement ?

a) Quelle réponse doit-on attendre à cette question ?
"OUI". Cela signifie que l'examen, c'est-à-dire l'ensemble des autres disciplines, sélectionne significativement les meilleurs parmi les bons.

b) Comment répondre à cette question ?
En comparant ainsi les cinq paires de moyennes :

$$\varepsilon = \frac{{}_5\bar{X} - \bar{X}^+}{\sqrt{S'^2_{\,5X} + S'^2_{\,X^+}}} \qquad \text{ou} \qquad \varepsilon = \frac{\bar{X}^- - {}_oX}{\sqrt{S'^2_{\,oX} + S'^2_{\,5X}}}$$

ε doit être positif et significativement différent de zéro ($\geqslant 1{,}96$)

c) Réponses :

	5C		oC	
	ε	α	ε	α
A	6,22	$< 10^{-9}$	1,1 NON	
BC	2,90	$< 10^{-2}$	1,1 NON	
BP	2,59	$\leqslant 10^{-2}$	0,95 NON	
H	5,001	$< 10^{-6}$	0,97 NON	
PHY	1,68 NON		3,72	$< 10^{-3}$

NON = ε non significatif

d) Conclusion :

L'examen sélectionne les meilleurs parmi les bons dans quatre disciplines : A - BC - BP - H ; il ne fait pas de différence parmi les bons en PHY. Il ne fait pas de différence parmi les mauvais, dans les mêmes quatre disciplines mais distingue les pires et les mauvais.

e) Discussion :

La réponse proposée comme "bonne", c'est-à-dire oui, est discutable. En effet, il n'était pas hors de mémoire ce fait que 5C représente seulement :
- 26 p.100 des acceptés (les bons en A)
- 32 p.100 des acceptés (les bons en BC)
- 35 p.100 des acceptés (les bons en BP)
- 27 p.100 des acceptés (les bons en H)

Sélection pour sélection, autant que ce fût la meilleure. La situation est assez différente en PHY : 5C représente 64 p.100 des bons. Il n'est donc pas scandaleux, le problème étant posé en d'autres termes, que des réponses opposées soient acceptées comme également satisfaisantes.

Il en va de même pour oC, mais en sens inverse : il y a peu de refusés en A - BC - BP et H et beaucoup en PHY. La sélection s'opère donc là où elle doit, et, surtout, là où elle peut.

QUATRIEME QUESTION

La population de 5C et celle de oC sont-elles homogènes ?

a) Quelle réponse doit-on attendre à cette question ?
"Oui". Pour que, enfin, il sorte de cet examen quelque chose qui ait l'air d'être indiscutable : que les reçus soient reçus, les refusés, refusés et ce, pour tout le monde.

b) Comment répondre à cette question ?
En comparant les variances ainsi :

$$F = \frac{S^2{}_{5X}}{S^2{}_{5Y}} \qquad \text{ou} \qquad F = \frac{S^2{}_{oX}}{S^2{}_{oY}}$$

Compte-tenu de l'effectif comparé (moins une unité dd1), F ne doit pas trop s'écarter de 1 pour qu'on puisse répondre "oui". Les dix rapports possibles seront envisagés.

c) Réponses :

	F de ^{5}C	F de oC
A.BC	NON	1,8
A.BP	NON	NON
A.H	NON	2,49
A	2,22	NON
BC.BP	NON	2,08
BC.H	NON	NON
BC	2,58	2,36
BP.H	NON	2,88
[illegible]	1,80	NON
H	2,83	3,26

Dans la mesure où la valeur de F est remplacée par NON cela signifie que la différence entre les variances n'est pas significative.

d) Conclusion :
Conclure est gênant. On peut cependant avancer que 5C est une classe plutôt homogène : l'accord est complet pour toutes les disciplines sauf PHY, avec qui le désaccord est complet. C^5 représente une certaine homogénéité. oC, en revanche, est plutôt hétérogène : une fois de

C I N Q U I E M E Q U E S T I O N

Les populations de $_5C$ et de oC sont-elles, respectivement, différentes de celles de ^1_4C ? En d'autres termes, les "certains" sont-ils bien différents des "incertains" ?

a) Quelle réponse doit-on espérer ?
"Oui", sans quoi la différence entre la certitude et l'incertitude est réduite à rien.

b) Comment répondre à la question ?
Comme précédemment, en comparant les variances, discipline par discipline, en 5C - ^1_4C d'une part, et oC-^1_4C d'autre part.

c) Réponses :

	F de 5C	F de oC
A	3,06	3,61
BC	2,39	4,39
BP	NON	NON
H	NON	2,79
PHY	NON	2,71

Dans la mesure où la valeur de F est remplacée par NON cela signifie que la différence entre les variances n'est pas significative.

d) Conclusions :
Il est utile d'examiner les disciplines une par une, en se référant pour chacune d'elles, aux réponses faites à la première question.

A établit une différence significative entre les moyennes, les rapports des différentes variances s'éloignant significativement de 1 : trois populations sont distinguées :
- les reçus,
- les incertains,
- les refusés.

BC fait les mêmes distinctions entre les trois mêmes catégories.

BP malgré la différence hautement significative de ces moyennes, considère tous les candidats comme appartenant à une même popu-

lation en considérant les catégories prises deux à deux. Cependant, le rapport des variances des acceptés (BP^+) et des refusés (BP^-) s'écarte significativement de 1. Donc BP ne considère pas que les (+) et les (-) appartiennent à la même population. Le partage se fait à l'intérieur de la catégorie des "incertains" qui apparaît ainsi, non pas admise, mais subie par cette discipline.

H et PHY de manière inattendue, présentent des caractères communs, malgré la différence des moyennes en 5C et ^1_4C, la population des reçus et celle des incertains est, en fait, la même, alors que celle des refusés et celle des incertains sont différentes.

R E S U M E

On pose à deux populations : oC et $_{5C}$ les questions suivantes :

1°) La moyenne des certains est-elle différente de celle des incertains ?

2°) Les moyennesdes certains diffèrent-elles entre elles ?

3°) Les moyennes des plutôt reçus est-elle différente de celle des franchement reçus et celle des plutôt refusés de celle des franchement refusés ?

4°) Les certains peuvent-ils se définir ?

5°) Les certains sont-ils différents des incertains ?

Les réponses ont été les suivantes, par comparaison avec la réponse attendue :

	Réponses attendues	Réponses faites en 5C	Réponses faites en oC
1°)	Oui	Oui 5/5	Oui 5/5
2°)	Non	Non 6/10	Oui 7/10
3°)	Oui	Oui 4/5	Non 4/5
4°)	Oui	Oui 6/10	Non 6/10
5°)	Oui	Non 3/5	Oui 4/5

Plus la question est précise, moins la réponse est franche. Ceux (5C) qui déclarent pouvoir définir les certains ne savent pas les distinguer des incertains, tandis que ceux qui se déclarent incapables de définir les certains les distinguent nettement des incertains ! Tout le monde s'accorde sur l'existence d'une différence entre la moyenne des certains et celle des incertains, mais beaucoup moins bien pour fixer cette moyenne, surtout chez les reçus. Tout diffère, rien n'est clair, beaucoup plus à cause de la manière de juger que de la façon d'exprimer son jugement.

CHAPITRE IV

GRANDEURS D'UN EXAMEN

> When you can measure what you are speaking about and express it in number, you know something about it.
>
> (Kelvin)

Il est constamment question de chiffres dans un examen. La plupart expriment des grandeurs. Si on accepte du terme "grandeur mesurable" la définition "terme pouvant entrer dans une des quatre opérations", bien peu sont des grandeurs mesurables.

INVENTAIRE DES GRANDEURS..... 70

SIGNIFICATION DES GRANDEURS..... 73

GRANDEURS MESURABLES OU BREVE HISTOIRE DES THERMOMETRES... 78

CONCLUSION..... 81

INVENTAIRE DES GRANDEURS

1° - L'effectif des reçus - Tout le monde y pense. Il s'exprime généralement en p.100 de l'effectif des candidats, parfois en pour 100 de l'effectif des étudiants. Confondre les deux expressions revient à dire que tous les candidats sont étudiants (ce qui est vrai) et que tous les étudiants sont candidats (ce qui est faux). C'est, souvent, le critère par excellence d'un examen :

- Un effectif reçu faible, est généralement perçu comme l'indice d'un examen difficile et "sérieux". Dans le cas contraire, on a plutôt tendance à penser que l'examen était "facile", trop facile, pas sérieux.

- A un effectif stable d'une session à l'autre, on attache généralement la signification d'une méthode stable : c'est alors un critère de qualité.

2° - La moyenne - C'est la note qu'auraient tous les candidats s'ils se partageaient également les points. Elle offre donc le reflet de ce que pense un correcteur de ses candidats : elle reflète la satisfaction du correcteur étant entendu qu'il manifeste une grande satisfaction par une note élevée, et une faible satisfaction, voire une insatisfaction, par une note basse. Evidemment, l'effectif des reçus dépend de la moyenne qui apparaît donc comme une grandeur plus importante puisqu'elle est déterminante, l'effectif étant déterminé.

Une moyenne élevée est prise comme un indice d'indulgence une moyenne basse traduisant une certaine sévérité. Il est souvent ressenti qu'une moyenne basse "fait sérieux", plus qu'une moyenne élevée. Mais une moyenne excessivement basse ferait excessivement sévère, une moyenne excessivement élevée ferait excessivement indulgent et, partant du principe que l'excès en tout est un défaut (jeu de mots d'ailleurs paradoxal), en ne corrigeant ni "trop haut", ni "trop bas" on obtient une moyenne "raisonnable" : proche de la médiane, avec laquelle on la confond souvent. A force d'être raisonnable, elle en devient inexpressive. La moyenne finalement, ne vaut que ce que vaut la correction.

3° - La variance - C'est une expression de la dispersion de la population autour de la moyenne. Elle mesure la variabilité et, plus précisément, la variété de la population. Il n'est pas inutile de se rappeler, concrètement, comment elle se calcule. Pour chaque candidat, on détermine l'écart qui le sépare de la moyenne : cet écart est donc un nombre de points. On porte cet écart au carré, on fait la somme des carrés de tous les écarts et on rapporte cette somme à l'individu moyen, en divisant la valeur obtenue par l'effectif (écart quadratique moyen). Le fait d'avoir pris le carré des écarts, joue, en quelque sorte, le rôle d'une loupe qui les grossit (mais tous avec la même méthode). La grandeur obtenue devrait s'exprimer, ici, en "points carrés". Comme ce n'est pas quelque chose de très concret, on préfèrera exprimer la grandeur en "points" en prenant la racine de la variance qui est l'écart type. Dans une distribution "normale" 95 p.100 de la population est éloignée de la moyenne en plus ou en moins, de deux écarts-types au maximum. La variance (ou son expression plus concrète : l'écart type) représente donc la variété de la population. Plus la variance est grande, plus la population est variée, plus elle est petite et plus la population est homogène. La variance, expression des écarts par rapport à la moyenne, ne vaut que ce que vaut la correction.

4° - La variance de la moyenne - Elle revêt une particulière importance, puisque la moyenne, c'est le critère le plus fameux, sinon le plus expressif. On peut lui donner plusieurs noms, dont deux sont

particulièrement significatifs :

- intervalle de confiance de la moyenne
- erreur standard.

Elle est directement liée à la variance et deux méthodes permettent de la calculer, chacune montrant à sa manière comment erreur stantard et variance sont liées.

1° - L'expression proprement variance, c'est-à-dire exprimée en "points carrés" est donnée par le rapport : variance de la moyenne = variance de la population divisée par l'effectif.

2° - L'expression commode, exprimée en points, appelée écart-type de la moyenne vaut :

a/ soit la racine de la variance de la moyenne (cette dernière étant le quotient de la variance par l'effectif de la population)

b/ soit l'écart-type lui-même (racine de la variance) divisé par la racine de l'effectif.

L'intervalle de confiance de la moyenne sera (au risque 5 p. 100 d'erreur), compris entre la moyenne augmentée ou diminuée de deux écarts types de la moyenne. Ainsi, l'intervalle de confiance de la moyenne, étroitement lié à la variance, dépend comme elle, de la correction.

Au total, toutes les grandeurs d'un examen dépendent de la correction. C'est peut-être une évidence, mais n'est-il pas plus sûr de la souligner ?

SIGNIFICATION DES GRANDEURS

1° - L'effectif des reçus

Il n'est que le reflet de toutes les opérations qui l'ont précédé. C'est un résultat, un signe, une conséquence. Sa connaissance n'apporte rien de particulier si on ne sait comment il a été obtenu. Cependant, c'est bien l'expression finale du résultat d'un examen. Pour un candidat, le résultat c'est de savoir s'il est placé dans l'effectif des reçus ou non. S'il s'est présenté, c'est dans l'espoir d'être dans le "bon" effectif. Pour l'enseignant, l'effectif c'est (ou ce devrait être) l'expression d'un résultat pédagogique. Si on exprime l'effectif des reçus en pour cent, c'est par rapport aux candidats qui se sont présentés. Ce pourcentage représente le rendement de l'enseignement. A supposer qu'on puisse comparer deux examens portant sur un même programme défini, que le premier se soit passé avant l'enseignement et le second après lui, on peut penser qu'au premier, personne ne serait reçu : l'enseignement manquait. Au second, on doit espérer que tout le monde serait reçu : l'enseignement a porté ses fruits. Le rendement est parfait on ne peut mieux demander. Loin d'être l'indice d'un mauvais examen, trop facile, une réussite parfaite (100 p. 100) devrait exprimer que l'enseignement a parfaitement atteint ses objectifs. Encore, faut-il, pour qu'une réussite parfaite indique un rendement parfait, qu'on puisse se fier au résultat. Il faut que l'examen soit fiable et

ce caractère dépend exclusivement de la compétence de l'examinateur. Dans ces conditions, et dans ces conditions seulement, on pourra tenter de partager équitablement les responsabilités qui reviennent pour de mauvais résultats à un mauvais enseignement et à de mauvais candidats.

Quelles que soient sa grandeur et sa stabilité, un effectif de reçus n'a de signification que si l'examen est fiable.

2° - La moyenne. C'est probablement la plus subjective des grandeurs examinées car elle est le souci constant du correcteur. C'est un fait, ce n'est pas une obligation. Pour celui qui examine a posteriori les résultats d'un examen, la moyenne est obtenue en faisant la somme de toutes les notes attribuées et en la divisant par l'effectif. Ceci est vraiment une moyenne. Pour le législateur, la moyenne c'est le point milieu qui sépare exactement la plus haute note possible (vingt en général) de la plus basse (zéro). C'est dix ; le onzième échelon d'une échelle qui en contient vingt et un. C'est une médiane. Nul ne peut être déclaré reçu s'il n'a obtenu au moins la moitié du maximum des points. Nul n'est censé ignorer la loi et personne n'ignore celle-ci : ni les étudiants, ni les correcteurs qui sont d'anciens étudiants.

Pour la plupart des correcteurs, la moyenne c'est une valeur définie a priori. Rien d'étonnant à ce que la moyenne constatée a posteriori s'écarte peu de la médiane et, bien souvent, de manière non significative. Qu'on en juge plutôt sur quelques chiffres empruntés à la session de juin 1971 :

A.	:	10,630	(63)
BC	:	10,246	(57)
BP	:	10,650	(65)
H	:	10,604	(72)
PHY	:	9,570	(41)

(Moyennes sur 20)

Ce qui n'empêche d'ailleurs pas que les effectifs acceptés soient différents d'une discipline à l'autre (valeurs entre parenthèses) : les écarts des effectifs sont sensiblement supérieurs à ceux des moyennes. Ainsi peut-on accuser les correcteurs de noter en fonction d'une valeur pré-établie. Le candidat semble-t-il bon ? Sa note aura deux chiffres, le second étant au moins "un".

Semble-t-il mauvais ? La note n'aura qu'un chiffre. Est-il bon, est-il mauvais ? Il aura dix. On peut en quelque sorte résumer la situation en disant que le devoir élémentaire d'un candidat est d'avoir la moyenne (et c'est d'ailleurs bien ce qu'il pense). Rien d'étonnant dans ces conditions, à ce que la moyenne observée soit la moyenne attendue. Pour que la moyenne ait une signification il serait préférable qu'elle ne soit pas le but et qu'elle ne se confonde pas avec la médiane.

3° - La variance. La moyenne ayant peu de signification, on peut attendre de la variance qu'elle apporte un peu de clarté. La moyenne prévue se confondra avec la moyenne observée ? Soit. Mais alors, les candidats s'en écartent, individuellement, plus ou moins. Autour de ce "zéro" imposé (la moyenne) on peut bâtir une dispersion en éloignant les franchement mauvais des franchement bons. Pour interpréter convenablement la variance d'un examen (ou tenter de le faire !) il faut tenter de distinguer ce qu'est un examen de ce qu'il devrait être.

Soit à mesurer le poids de cent personnes pesant chacune soixante dix kilos.

a) Un même observateur va peser chacune de ces personnes sur plusieurs balances. Si chaque personne n'accuse pas le même poids sur chaque balance, son poids va apparaître comme variable. Ainsi un facteur balance sera mis en évidence et ce facteur sera exprimé pour chaque balance et pour l'ensemble de la population, par la variance du poids.

b) Plusieurs observateurs vont peser chacune de ces personnes sur la même balance. Si chaque personne n'accuse pas le même poids, un facteur observateur sera mis en évidence et exprimé, pour chaque observateur, par la variance du poids.

Si les deux opérations sont diversement combinées, on pourra ou non mettre en évidence un facteur observateur et un facteur balance, mais on pourra toujours mettre en évidence un facteur mesure puisqu'on sait, et qu'on tient pour acquit, que chaque personne pèse un poids déterminé et, naturellement, fixe.

Peser chacune des personnes sur la même balance, cela

revient à poser plusieurs fois la même question (ou les mêmes questions) aux candidats. Utiliser la même balance et plusieurs observateurs, c'est la double correction.

Soit à mesurer la taille de cent individus. Chacun sera mesuré plusieurs fois, avec une méthode chaque fois différente, les résultats étant portés sur un tableau récapitulatif correspondant à une série de mesures.

La première mesure se fera avec un bâton vertical comportant un unique repère : 1,70 m. Si tout l'effectif de l'échantillon mesure "à peu près" 1,70 m, chacun sera déclaré mesurer 1,70 m.

La seconde se fera avec une toise graduée en décimètres. Dans les conditions précédentes chacun sera déclaré "mesurer 17 décimètres" : la dispersion est nulle.

La troisième se fera avec une toise graduée en centimètres : on pourra distinguer 170 cm de 171 ou de 169 : une dispersion apparaît.

La quatrième avec une toise graduée en millimètres, permettra de distinguer 1700 millimètres de 1 701 ou de 1 699 millimètres : la dispersion est plus grande. Ainsi, plus précise sera la mesure, plus grande sera la variance. Inversement, on peut déduire que plus la variance est grande, plus les mesures sont précises.

En est-il de même pour un examen ? Certainement non. En effet, mesurer des poids, ou des longueurs, c'est les comparer à un étalon. Or, il n'y a pas d'étalon défini dans un examen. On ne peut donc pas parler de "précision". En revanche, on peut prévoir que plus la note est discriminative et plus la variance sera grande. En terme d'examen la variance traduit non pas la précision de la mesure, mais sa sensibilité. D'autre part, lorsque plusieurs mesures sont faites (plusieurs épreuves) et quelle que soit la moyenne observée, on peut affirmer que si la variance diffère, la méthode d'évaluation n'est pas la même si elle s'adresse à une même population d'individus car cela signifie que la variété estimée de la population n'est pas la même. C'est l'estimation qui diffère, pas la population. Une différence de variance met en évidence un facteur examen (et non un facteur candidat) sans qu'il soit possible de discerner sur ces seules données la nature du facteur examen.

4° - La variance de la moyenne

Supposons résolues les difficultés de la correction : pas d'erreur de mesure, un étalon est défini, les mesures faites sont justes, sensibles et fidèles. Tout est pour le mieux dans le meilleur des mondes qu'examinateur puisse souhaiter. Il n'est plus question de noter par rapport à une moyenne subie (car elle n'est pas imposée). La moyenne observée a donc une signification. Mais elle n'est pas exacte. En effet, et quelle que soit la répartition des notes : normale (gaussienne) ou non, pour peu que la population observée soit assez nombreuse (et c'est le cas dans un examen concernant une année de Faculté) la loi de probabilité de la moyenne est "normale". Cela revient à dire que la moyenne fluctue, au risque 5 p. 100, entre la moyenne observée et plus ou moins deux écarts types de la moyenne.

Si la population n'est pas constituée de jumeaux (ce qui est hautement probable !) elle présente dans les caractéristiques mesurées, une certaine dispersion exprimée par sa variance. Toutes choses égales par ailleurs, plus la mesure sera sensible, plus la variance des individus apparaîtra grande. Donc, plus grande sera la variance de la moyenne, qui est liée à la précédente par le facteur 1/n. Plus grande est la variance de la moyenne, plus grand est l'écart type (c'est sa racine carrée). Plus grand est l'écart type de la moyenne, plus large est l'intervalle de confiance de la moyenne. Plus large est l'intervalle de confiance, moins précise est la moyenne.

Ainsi "meilleure" serait la correction, moins précise serait la moyenne, ce qui semble-t-il n'est pas une évidence pour la plupart des examinateurs qui accordent à une moyenne, artificielle en fait, une confiance ridicule. La conséquence, c'est qu'un certain nombre de candidats qui s'écartent de la moyenne, du mauvais côté, mais non significativement, sont refusés (malchance !). Il est vrai qu'un certain nombre d'autres qui, eux, s'écartent du bon côté mais non significativement sont reçus (chance !). Un certain nombre ? Il n'est que de regarder certains des histogrammes déjà examinés pour se rendre compte que ce nombre est parfois loin d'être négligeable !

GRANDEURS MESURABLES OU BREVE HISTOIRE DES THERMOMETRES

La citation préliminaire de Lord KELVIN mérite réflexion. "Lorsque vous pouvez mesurer ce dont vous parlez et l'exprimer par un nombre, vous en savez quelque chose". Il n'y est pas seulement question d'exprimer ce dont on parle par des chiffres mais de pouvoir préalablement mesurer. Autre chose, en réalité, amplifie beaucoup le champ de réflexion, suggéré par le nom même de l'auteur : l'expression des températures. Deux températures ne peuvent pas s'additionner : la température n'est pas une grandeur mesurable. Deux températures peuvent se comparer pour peu qu'elles soient exprimées dans le même système. Mais elles ne peuvent pas entrer dans la plus simple des opérations : l'égalité. Soit à comparer 37°C et 38°C. Tout ce qu'on peut en dire c'est qu'une température est plus élevée que l'autre, que l'écart est de un degré centigrade entre elles (ou 1° K) que l'une est plus éloignée que l'autre d'une origine choisie, que ces températures présentent le même écart absolu qu'on l'exprime en "degrés KELVIN" ou en "degrés CELSIUS" mais qu'elles diffèrent l'une de l'autre relativement plus si on les exprime en Celsius que si on les exprime en Kelvin. A plus d'un titre, la comparaison s'impose entre "noter un examen" et "prendre une température". D'abord, et peut être surtout, dans l'histoire de la thermométrie : on y retrouve la recherche d'un étalon

Le premier thermomètre méritant citation est celui de Cornelius VAN DREBBEL (1572-1632). Ce Hollandais avait inventé un thermomètre à gaz qui fournissait des chiffres mais, comme il était sensible et à la température et à la pression, on ne pouvait jamais très bien discerner ce que mesurait cet instrument : il n'était pas sélectif.

L'Accademia del Cimento (à Florence) construisit, au XVI° siècle, un thermomètre à alcool qui n'était sensible qu'à la température. Restait à l'étalonner. Les académiciens recherchèrent ce qu'il y avait de plus froid chez eux et ce fut une certaine cave : cette température s'appela zéro. La colonne d'alcool devait s'arrêter à la température la plus élevée qui fut choisie comme celle régnant à midi au jour le plus chaud de l'été à Florence. Le système, préférable au précédent, manquait d'universalité.

L'Irlandais Robert BOYLE (1626-1691) proposa un zéro vraiment universel : la température de congélation de l'eau. Ce n'est pas pour autant qu'une échelle fut définie. Tandis que les Britanniques de la Royal Society établissaient leur propre échelle de 65° (température de congélation de l'eau) à zéro degré (qui était une température beaucoup plus élevée !) que certains physiciens étalonnaient leurs thermomètres sur les fébricitants (!), Daniel FAHRENHEIT définit l'échelle qui porte son nom en 1714. Il avait retenu trois choses pour cela, dont deux "objectives" :

1° - ayant pris connaissance de la découverte qu'avait fait en 1703 le physicien français Guillaume AMONTONS : l'eau bout à température fixe, il appela 212ème degré cette température.

2° - la température de congélation de l'eau fut le 32ème degré.

3° - le zéro, la référence, fut fourni par un certain mélange réfrigérant de l'invention du physicien prussien.

Aux environs de 1710, REAUMUR, fixa le zéro là où il est resté depuis. Sa seconde référence fut la température d'ébullition de l'eau mais, malheureusement, l'échelle était octogésimale. Aussi l'échelle que proposa Anders CELSIUS en 1742 fût-elle la bienvenue pour sa simplicité : elle était centésimale et fut adoptée dans presque tous les pays. Ainsi, plus de cent ans de

tâtonnements furent-ils nécessaires pour définir un étalon et une échelle. En 1851 (141 ans de plus !) THOMSON, qui n'était pas encore Lord KELVIN, déduisit de la loi de Gay LUSSAC, la situation du zéro "absolu".

En gros, et en ce qui concerne les examens, nous en sommes plutôt à étalonner nos températures sur celles des fébricitants. Ce serait un progrès que de parvenir à 1710 et notre échelle vigintésimale n'aurait plus qu'à attendre son CELSIUS.

Par sa nature même, l'examen est une comparaison. Une comparaison entre ce qu'est l'étudiant jugé et ce qu'on pense qu'il devrait être au moins. C'est une comparaison entre un sujet défini et un étalon ou, plus précisément, une référence non définie. Non définie, mais non pas inconnue car, pour une large part, c'est l'image que se fait l'examinateur du "candidat moyen" qui sert de référence. Il n'est pas pour autant certain que cette image puisse être reproduite, puisqu'elle n'est, en aucune façon, définie. Ainsi deux défauts entachent l'évaluation des candidats

- elle est fondée sur l'image que se fait l'examinateur du candidat moyen : c'est subjectif
- cette image se traduit avec une assez grande fidélité, par la même note : la moyenne, même si la valeur du candidat moyen fluctue en fonction du temps ou des épreuves : c'est arbitraire.

CONCLUSION

1° - Les effectifs ne s'ajoutent pas.

Dans une épreuve ils peuvent s'ajouter : cinq candidats reçus, c'est deux fois moins que dix ou cinq fois plus qu'un. Doubler le nombre des candidats reçus c'est bien multiplier par deux un effectif et obtenir un effectif final deux fois plus grand. Mais accepter dix candidats de plus dans une épreuve ce n'est pas nécessairement recevoir dix candidats de plus à l'examen. Dans l'examen qui a fourni la plupart des chiffres (juin 1970) l'effectif reçu à l'examen était de 142. Aux différentes épreuves il y en avait 207, 171, 156, 299, 85. Ce n'est certainement pas en faisant la somme de ces effectifs élémentaires qu'on peut trouver 142. Les effectifs d'une épreuve sont une grandeur mesurable. Les effectifs reçus à l'examen, NON.

2° - Les moyennes ne s'ajoutent pas

Ce sont des valeurs fictives. C'est la valeur qu'auraient tous les individus s'ils se la partageaient également. Ce sont des grandeurs de comparaison. Ce ne sont pas des grandeurs mesurables. La note d'un individu, c'est une moyenne : la moyenne de la valeur attribuée à chaque réponse faite, chacune affectée d'un coefficient peu définissable censée représenter "l'importance". C'est le reflet simple d'un tout complexe. Si la note a les dimensions d'une moyenne, que les moyennes ne s'ajoutent pas, les notes ne s'ajoutent pas. Ne devraient pas s'ajouter.

3° - Les variances ne s'ajoutent pas

Ce sont des comparaisons par nature : elles expriment des écarts par rapport à une référence définissable : la moyenne. C'est d'autre part une moyenne (écart quadratique moyen). Ce qui suffit à exprimer le fait qu'il ne s'agit pas de grandeurs mesurables.

Au total, aucune des grandeurs servant à exprimer les résultats d'un examen, ne sont des grandeurs mesurables. Aucune ne peut, entrant dans une addition, fournir une somme significative. Or, le résultat d'un examen, c'est une somme de notes.

CHAPITRE V

GRANDEURS COMPARABLES ?

On ne peut comparer que ce qui est comparable.

LA PALISSE

LES EFFECTIFS..... 87

LES MOYENNES..... 91

EFFECTIFS ET MOYENNES..... 94

LES VARIANCES..... 98

Si le résultat d'un examen n'est pas une somme mais une moyenne (ce qui techniquement revient au même) cette moyenne, celle du candidat, doit être comparée avec une moyenne de référence : celle qui définira le candidat "suffisant". Dans la mesure, et c'est ici la règle, où un examen comporte plusieurs épreuves, il faut, pour déclarer "suffisant" un candidat, le comparer, pour chacune des épreuves, à la référence de chacune d'entre elles. On aboutira à un ensemble de comparaisons qui permettront de dire que le candidat jugé est suffisant ici et insuffisant là, ou suffisant partout ou nulle part. S'il est déclaré suffisant partout ou nulle part, le jugement est simple : il est finalement "reçu" ou "refusé". Mais ce cas simple n'est pas le plus fréquent (29 p. 100 en 1970, 34 p. 100 en 1971, différence non significative). Dans 70 p. 100 environ des cas, le jugement final dépendra de comparaisons : de la "moyenne" du candidat. Le bien fondé de cette habitude (est-ce une méthode ?) mérite analyse.

Que les grandeurs qui permettent de caractériser les résultats d'un examen soient différentes d'une épreuve à l'autre n'est pas un facteur de scandale. Les épreuves sont différentes, elles ne font pas appel aux mêmes connaissances. Pour des motifs divers (aptitude, capacité, attirance....) un candidat peut parfaitement être "meilleur" ici que là. Et, même, "bon" ici, "mauvais" là. Le jugement exprimé n'est pas nécessairement faux parce qu'il est différent. Inversement, l'identité de deux jugements n'est pas une preuve de leur justesse. Sur les données d'un seul examen, on ne

peut que constater que des jugements différents ont été émis par cinq épreuves de nature différente. Impossible de savoir qui a raison ou tort. D'une année à l'autre, les candidats ont changé. Pas les disciplines. Peut-être est-il possible d'établir un rapport entre des fluctuations de méthode et des fluctuations de résultats en ne perdant pas de vue que :

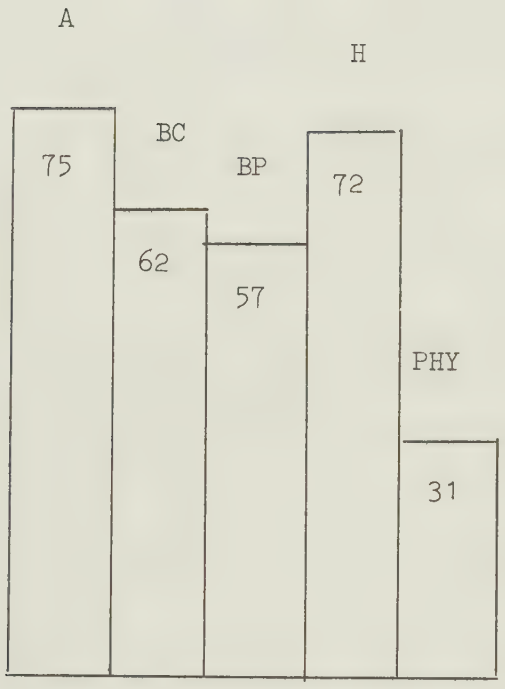
1° - dans une même année, cinq épreuves différentes jugent les mêmes individus,

2° - d'une année à l'autre, chaque épreuve juge une population différente.

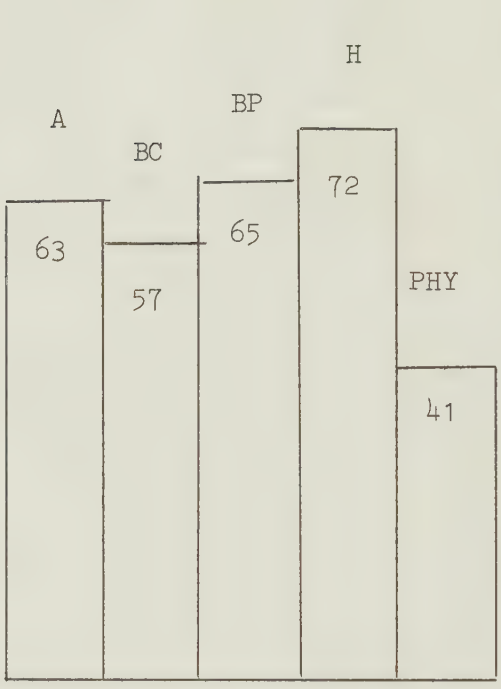
Si des méthodes différentes apparaissent, les résultats ne seront pas comparables. Trois tests seront successivement appliqués : comparaison des effectifs, des moyennes et des variances.

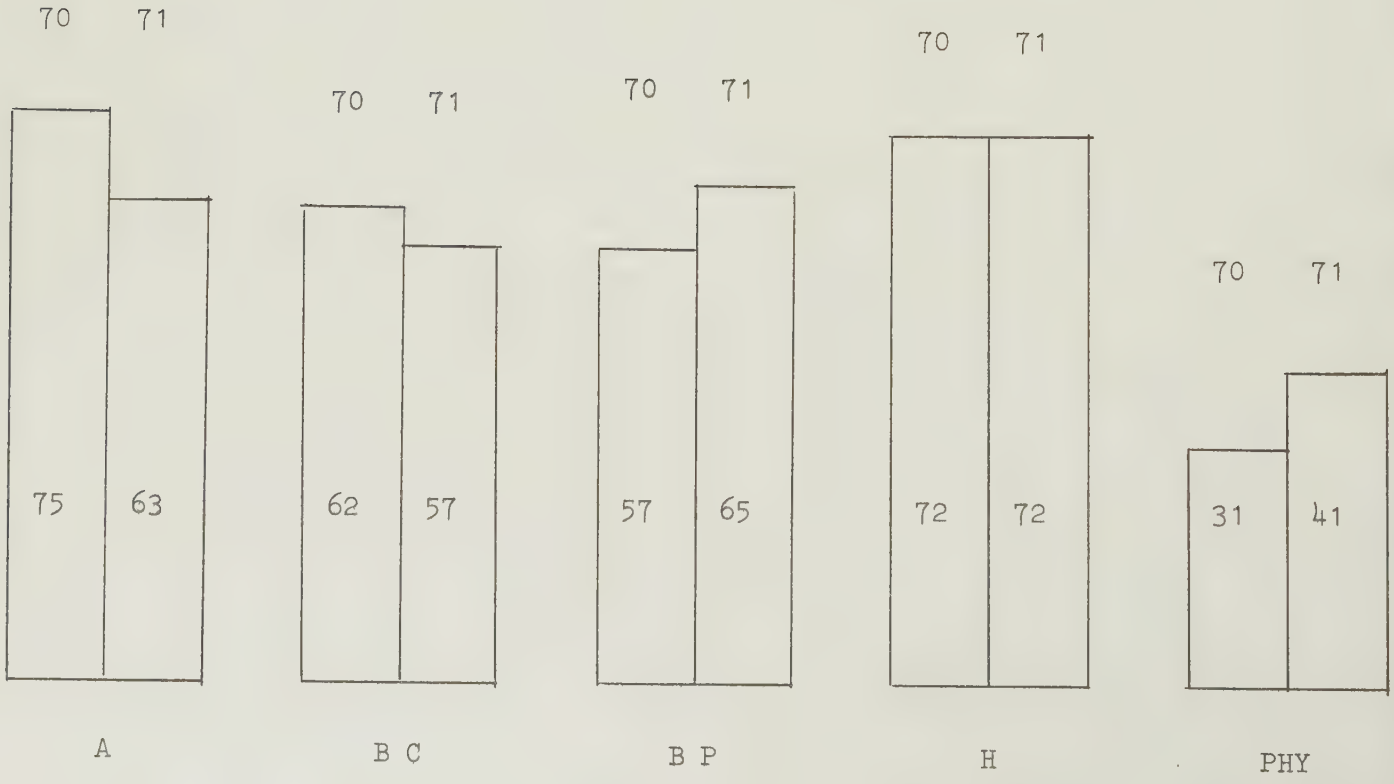
Comparaison des Effectifs Reçus.

1970



1971





LES EFFECTIFS

A - en 1970

a) Ils sont différents.

Ceci peut signifier deux choses :

. Les corrections sont inégalement dures (facteur examen)

. Les candidats préfèrent certaines disciplines (ou les réussissent mieux : facteur candidat)

Le choix est impossible entre les deux explications.

b) Ils sont répartis dans un certain ordre décroissant

A, H, BC, BP, PHY

c) Ils sont significativement différents : pour 8 couples sur 10.

Exceptés les couples : A.H

BC.BP

B - en 1971

a) Ils sont différents les uns des autres

b) Ils sont répartis dans un certain ordre décroissant

H, BP, A, BC, PHY

c) Six couples sur dix diffèrent significativement. Quatre ne diffèrent pas :

- A.BC
- A.BP
- BP.H
- BC.BP

C - Comparaisons

a) des similitudes

1° - Les effectifs sont toujours différents

2° - Les couples liés sont formés de disciplines de rang voisin : 1ère et 2ème, 3ème et 4ème et même 2ème et 3ème ce qui n'était pas très clair en 1970 mais net en 1971

3° - PHY reste en dernière position, significativement écartée de toutes les autres

4° - BC et BP restent liées.

b) des différences

1° - L'ordre décroissant des effectifs a changé

2° - A et H ne sont plus liées en 1971

3° - A.BC, A.BP, BP.H sont liées en 1971 et ne l'étaient pas en 1970

4° - Les effectifs diffèrent moins les uns des autres en 1970 qu'en 1971.

D - Interprétation

Un pourcentage n'a pas changé du tout : celui de H. Deux autres n'ont pas changé significativement : celui de BC (qui a diminué) et celui de BP (qui a augmenté). Cela suffit déjà pour dire que les changements de rang n'ont pas grande signification.

Le pourcentage de A a significativement diminué (de 75 à 63) et celui de PHY significativement augmenté (de 31 à 41).

Ainsi, aux yeux des cinq disciplines, la population de 71, comparée à celle de 70

- n'a pas changé : BC, BP, H

- s'est améliorée : PHY (BC a tendance à partager cet avis)
- est moins bonne : A (BP a tendance à partager cet avis)

En ne tenant pas compte des différences non significatives, la situation peut être ainsi résumée :

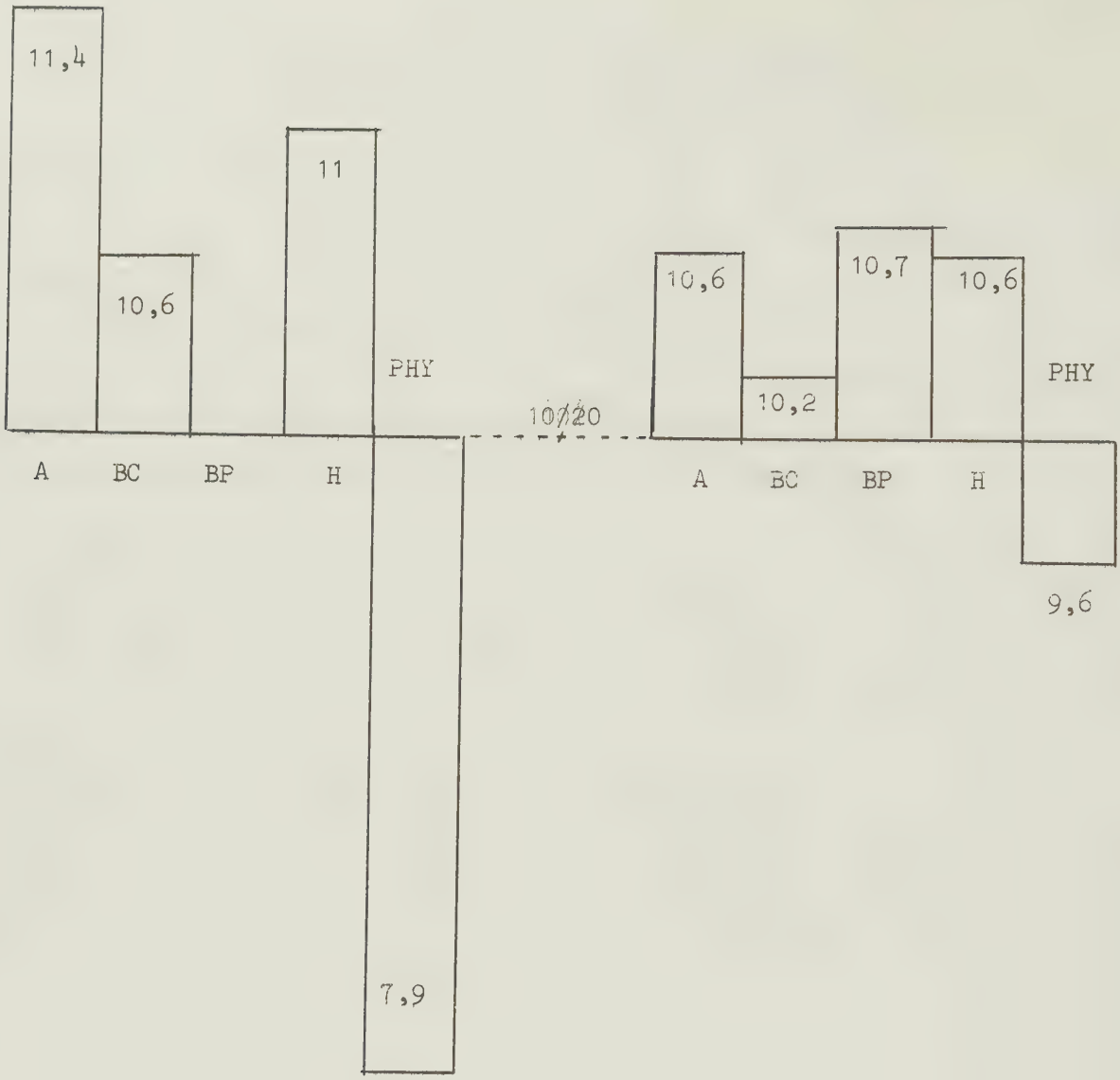
1° - Pour A, ou bien la correction s'est durcie (facteur examen) ou bien les candidats sont réellement moins bons (facteur candidat).

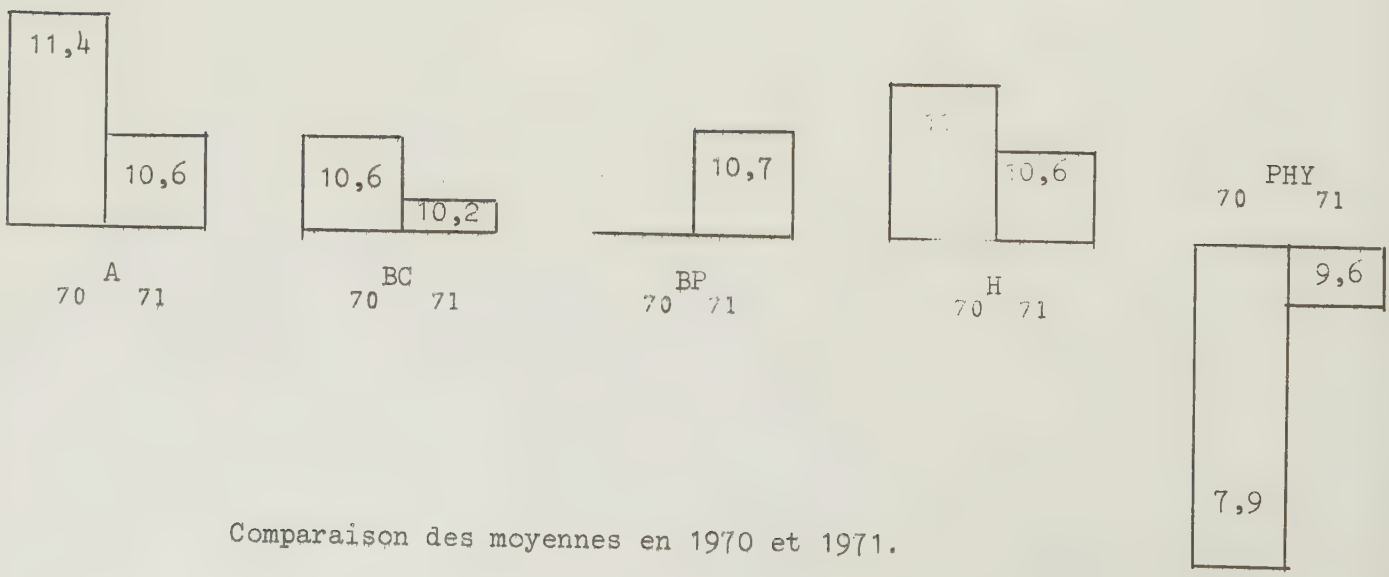
2° - Pour PHY, ou bien la correction s'est adoucie (facteur examen) ou bien les candidats sont réellement meilleurs (facteur candidat).

3° - Pour BC, BP et H les hypothèses précédentes peuvent être retenues : meilleurs candidats plus sévèrement jugés, moins bons candidats moins sévèrement jugés, candidats identiques jugés de manière identique.

Faute de connaître un reflet fidèle de la correction, force est d'admettre que ces fluctuations d'effectifs n'ont pas de signification définissable. La comparaison de leur grandeur n'a pas de signification.

1970 1971





Comparaison des moyennes en 1970 et 1971.

LES MOYENNES

A - en 1970

a) Elles sont différentes

b) Elles sont réparties dans un certain ordre décroissant :
A, H, BC, BP, PHY qui est celui des effectifs

c) Elles sont significativement différentes sauf dans les couples :
A.H
BC.BP
BC.H

B - en 1971

a) Elles sont différentes

b) Elles sont réparties dans un ordre décroissant :
BP, H, A, BC, PHY : différent de celui des effectifs

c) Elles ne diffèrent pas significativement dans six couples sur dix :
- A.BC
- A.PHY
- BC.BP
- BC.H
- BP.PHY
- H.PHY

C - Comparaisons

a) des similitudes

1° - les moyennes diffèrent toujours dans les couples suivants (3 sur 10)

- A.BP
- BC.PHY
- BP.H

2° - les moyennes restent sensiblement égales entre elles dans les couples suivants (2 sur 10)

- BC.BP
- BC.H

3° - la moyenne de PHY reste la plus basse.

b) des différences

1° - l'ordre décroissant des moyennes a changé

2° - l'ordre des moyennes ne correspond pas en 71 à celui des effectifs

3° - les moyennes se rapprochent les unes des autres : six des dix comparaisons possibles ne montrent plus d'écart significatif.

D - Interprétation

Trois moyennes n'ont pas sensiblement varié (différences non significatives) celles de :

- A : qui a baissé
- BC : qui a baissé
- H : qui a baissé

Deux ont significativement augmenté : celle de BP et celle de PHY.

On peut ainsi résumer la situation :

1° - Pour BP et PHY : ou bien la correction s'est adoucie ou bien les candidats sont meilleurs.

2° - Pour A, BC et H, les candidats sont moins bons mais moins sévèrement jugés ou meilleurs et plus sévèrement jugés ou identiques et identiquement jugés.

Si on compare la sévérité des corrections en prenant comme critère successivement les effectifs et les moyennes on obtient les résultats suivants (70 étant la référence).

	EFFECTIFS	MOYENNES
A	↗	=
BC	=	=
BP	=	↘
H	=	=
PHY	↘	↘

Le graphique représente la "sévérité", les candidats étant supposés inchangés dans leur valeur.

On voit que les avis ne concordent que deux fois (BC et H). Si on suppose que chaque épreuve juge de façon stable d'une année à l'autre, son jugement s'exprime deux fois seulement de façon convergente. Il semble bien que moyennes et effectifs ne soient pas deux grandeurs comparables.

EFFECTIFS ET MOYENNES

Le tableau suivant résume la situation. Il comporte en plus la valeur des effectifs et des moyennes.

	EFFECTIFS P. 100			MOYENNES SUR 20		
	70		71	70		71
A	75	↘	63	11,3	=	10,68
BC	62	=	57	10,22	=	9,91
BP	57	=	65	9,88	↗	10,72
H	72	=	72	10,86	=	10,69
PHY	31	↗	41	7,92	↗	9,86

En BC et H, les moyennes n'ont pas sensiblement varié, les effectifs reçus non plus. Ou bien l'affinité des candidats pour ces disciplines n'a pas changé ; les résultats sont logiques. Ou bien les candidats ont changé et ces disciplines ne l'ont pas perçu.

En PHY la moyenne a augmenté, l'effectif reçu aussi. C'est logique ; mais qui a changé ? La correction qui s'est adoucie (facteur examen) ou les candidats qui se sont améliorés (facteur candidat) ? + 9,7 p. 100 de moyenne donne + 10 p. 100 d'effectif.

En A, la moyenne n'a pas significativement changé alors que l'effectif des reçus a significativement diminué. Il est vrai que la moyenne a en réalité baissé. L'effectif a varié dans le même sens. Une diminution de la moyenne de 0,62 sur 20 (3 p. 100) a entraîné une diminution de l'effectif reçu de 12 p. 100. Une variation insensible de la moyenne a sensiblement fait varier l'effectif : un facteur examen peut être soupçonné.

En BP : la moyenne a augmenté, l'effectif aussi mais non significativement. Une augmentation de 4,2 p. 100 de la moyenne augmente l'effectif de 8 p. 100.

Moyennes et effectifs ont dans l'ensemble varié et varient dans le même sens. Mais un rapprochement reste intéressant : la même variation de moyenne (m) entraine-t-elle la même variation d'effectif (n) ?

A	:	m	- 3	p.100	n	- 12 p.100
BC	:	m	- 1,55	p.100	n	- 5 p.100
BP	:	m	+ 4,2	p.100	n	+ 8 p.100
H	:	m	- 0,85	p.100	n	$\pm$ 0 p.100
PHY	:	m	+ 9,7	p.100	n	+ 10 p.100

Pour A et BC les chutes d'effectif sont à peu près proportionnelles (et proportionnellement égales) aux chutes de moyenne.

En BP, une amélioration de la moyenne supérieure à la variation de celle de A entraîne une variation d'effectifs inférieure à celle de A. Une amélioration deux fois plus petite de la moyenne qu'en PHY entraîne une augmentation des effectifs presque égale. Les variations de H restent négligeables.

En PHY, outre la comparaison déjà faite avec BP, on peut remarquer que les variations de moyenne entraînent des variations d'effectif tout à fait différentes de celles de A et BC. Une variation de moyenne en A qui n'est pas le tiers de celle de PHY entraîne une variation d'effectif supérieure tandis qu'une variation moyenne

de BC six fois plus petite que celle de PHY entraîne une variation d'effectif seulement deux fois plus petite.

Au total, il semble bien qu'un facteur examen existe.

	A	BC	BP	H	PHY
Variances des effectifs 70	0.001.875	0.002.356	0.002.451	0.002.016	0.002.139
Variances des effectifs 71	0.002.331	0.002.451	0.002.275	0.002.016	0.002.419
Variances des notes 70	25.41	12.53	8.87	7.02	16.73
Variances des notes 71	19.94	9.54	7.24	9.23	8.92
ε de S^2 70 - S^2 71 (notes)	2.7633	3.1061	2.3251	3.0838	6.7783
F des variances des effectifs	1.24	1.04	1.07	1.00	1.13
F des variances des notes	1.27	1.31	1.22	1.31	1.87

ε = écart réduit
S^2 = variance

$$F = \frac{S^2\ b}{S^2\ b}$$

COMPARAISON DES VARIANCES DES DIX COUPLES.

(VALEURS DE F)

	RECUS	NON RECUS	TOTAL
A BC	1.99	1.99	2.02
A BP	1.46	3.76	2.86
A H	1.76	5.44	3.61
A PHY	1.21	1.92	1.51
BC BP	1.36	1.88	1.41
BC H	1.13	2.72	1.78
BC PHY	2.43	1.03	1.33
BP H	1.20	1.44	1.26
BP PHY	1.77	1.95	1.88
H PHY	2.14	2.82	2.38

LES VARIANCES

En 1970, elles varient toutes, significativement, les unes des autres. Y compris :

- celles de A et H qui avaient des effectifs comparables et des moyennes comparables
- celles de BC et de BP qui avaient des effectifs comparables et des moyennes comparables.

Par conséquent, et pour ces deux couples de disciplines, les populations ont été jugées identiques par leur valeur et leur nombre mais pas pour leur variété. Malgré les apparences, A et H aussi bien que BC et BP ont jugé comme différentes des populations qui se trouvaient avoir même moyenne et même pourcentage de réussite. Ainsi est démasqué le facteur examen : une même population, composée des mêmes individus est assurément jugée de manière différente, même par des disciplines qui donnent la même note et fournissent le même nombre de reçus. Toutes les disciplines ne perçoivent pas également la variété de la population. Les différences de moyennes observées peuvent être le fait d'un facteur candidat. Pas le jugement sur la variété.

En 1971, elles diffèrent toutes significativement quoiqu'à un moindre degré. Mais ce qui est intéressant, c'est de les comparer, discipline par discipline, année par année. Le tableau joint résume ces variations (70 étant la référence) en précise le sens, mais pour plus de clarté non la valeur et, dans un synopsis, permet les rapprochements avec moyennes et pourcentages.

	EFFECTIFS	TOUS candidats		candidats RECUS		cand.NON RECUS	
		moyennes	variances	moyennes	variances	moyennes	variances
A	↘	↘	↘	=	↘	↘	↘
BC	= (-)	=	↘	=	↗	=	↘
BP	= (+)	↗	↘	=	=	↗	↘
H	=	=	↗	=	↘	↘	↗
PHY.	↗	↗	↘	=	↘	↗	↘

Une première discipline sera analysée : BC. Le pourcentage d'acceptés en 1971 ne diffère pas significativement de celui de 1970, non plus que la moyenne des candidats. Allant plus loin, on peut constater que la moyenne des candidats acceptés reste la même ainsi que celle des candidats refusés. Voilà, peut-on penser, un modèle de stabilité, de fidélité. Voire ! La variance des notes a changé : elle s'est abaissée. On peut donc penser que la discrimination, la sensibilité de la correction, est moindre en 1971. Le calcul de l'écart réduit montre que cet abaissement est significatif ($\mathcal{E}$ = 3,1061). La différence n'a pas une chance sur cent d'être sans signification. L'hypothèse est donc : en 1971, BC a corrigé de façon moins précise. Pour le démontrer, il faut trouver de l'imprécision quelque part. Les acceptés ont une moyenne stable, d'une année à l'autre, mais une variance augmentée : donc, en suivant logiquement ce jeu d'hypothèses, on peut en déduire que les acceptés ont été corrigés de façon plus sensible. On pourrait même oser tenter de dire qu'ils sont plus franchement acceptés en 1971. La variance des notes des candidats étant plus basse, celle des acceptés étant plus haute, rien d'étonnant à ce que la variance des refusés ait baissé. Au total, on peut poser la conclusion suivante :

- BC a corrigé de façon moins sensible en 1971
- Les acceptés sont appréciés de façon plus sensible (donc plus précise)
- Les refusés sont appréciés de façon moins sensible (donc moins précise).

L'incertitude a augmenté en ce qui concerne les refusés. Reste à le prouver. Pour cela, recherchons l'incertitude là où elle est : la moyenne des non reçus, augmentée de deux écarts types donne la limite <u>supérieure</u> (au risque 5 p. 100) de la note des non reçus. La moyenne des reçus, diminuée de deux écarts types, donne la limite <u>inférieure</u> (au risque 5 p. 100) de la note des reçus. La zone de recouvrement, c'est la zone d'incertitude.

La moyenne des non reçus et les écarts types correspondant ont été successivement :

- 6,59/20 et 7,15/20
- 2,35 et 1,74 (les écarts types ont naturellement varié dans le même sens que les variances).

La note limite supérieure des non reçus en 1970 et 1971

est respectivement :

- 6,59 + 2 x 2,35 = 6,59 + 4,70 = 11,29
- 7,15 + 2 x 1,74 = 7,15 + 3,48 = 10,63

On voit donc que les acceptés le sont certainement au-dessus de 11 tant en 1970 qu'en 1971. Compte tenu du fait que cette discipline note au quart de point, on peut même préciser que :

- en 1970 on est accepté (risque 5 p.100) à 11,5/20
- en 1971 on est accepté (risque 5 p.100) à 10,75/20

La zone d'incertitude, du côté des acceptés, s'est bien rétrécie.

La moyenne des reçus et les écarts types correspondants ont été successivement :

- 12,46 et 12,23
- 1,76 et 2,12

Les notes limites inférieures des reçus en 1970 et 1971 sont respectivement, au risque 5 p.100 :

12,46 - 2 x 1,76 = 12,46 - 3,52 = 8,94
12,23 - 2 x 2,12 = 12,23 - 4,24 = 7,99

On voit que les refusés le sont certainement au-dessous de 9 en 1970 et au-dessous de 8 en 1971. La zone d'incertitude, du côté des refusés s'est bien élargie.

Au total, moins d'acceptés risquent de l'avoir été par chance, mais plus de refusés risquent de l'avoir été par malchance. Une fois de plus on constate que l'acceptation est mieux définie que le refus et, pour cette discipline, ce défaut s'est aggravé.

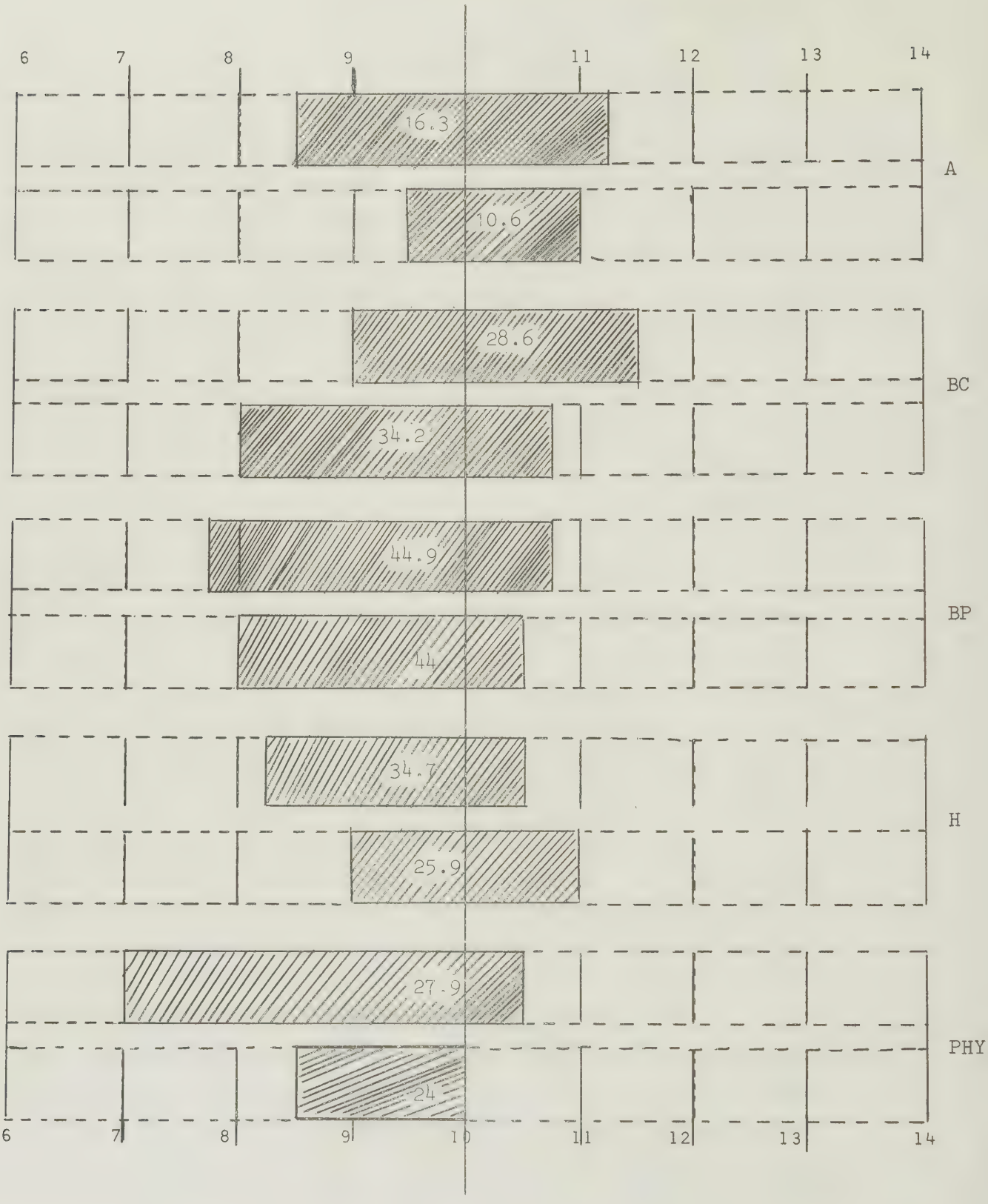
La même opération peut être faite en H : les données concernant pourcentage et moyenne sont sensiblement les mêmes, inverses pour les variances. On peut en déduire que les acceptés le sont avec moins de précision et les refusés avec plus de précision.

1970 : refusés : 7,65 + 2 x 1,42 = 7,65 + 2,84 = 10,49
acceptés : 12,08 - 2 x 1,87 = 12,08 - 3,74 = 8,34

1971 : refusés : 6,87 + 2 x 2,06 = 6,87 + 4,12 = 10,99
acceptés : 12,26 - 2 x 1,66 = 12,26 - 3,32 = 8,94

Conformément à ce qu'avaient laissé prévoir les différences

VARIATIONS DES ZONES D'INCERTITUDE



Les rectangles figurent les zones d'incertitude

Dans chacun d'eux,le pourcentage d'effectif "incertain" est précisé

de variances :

- au total : l'incertitude a diminué (de 2,15 à 2,05)
- chez les acceptés, elle a augmenté ; la note de réception "sûre" s'est élevée (rapprochée du maximum)
- chez les refusés, elle a diminué : la note de refus "sûr" s'est élévée (éloignée du minimum).

L'expression choisie : rapprochée ou éloignée du maximum ou du minimum a l'avantage de faire disparaître le mot de moyenne dont la valeur n'est pas nécessairement la même d'une discipline à l'autre et, dans une même discipline, d'une année à l'autre.

Les trois disciplines qui restent ont toutes diminué leurs variances. Leur incertitude a donc dû augmenter. Cela peut se vérifier au niveau de PHY :

- 1970 : refusés : $5,66 + 2 \times 2,24 = 5,66 + 4,8 = 10,46$
  acceptés : $12,52 - 2 \times 2,74 = 12,52 - 5,48 = 7,04$
- 1971 : refusés : $7,66 + 2 \times 1,1 = 7,66 + 2,2 = 9,86$
  acceptés : $12,78 - 2 \times 2,03 = 12,78 - 4,06 = 8,72$

En 1970 : on était reçu au-dessus de 10,46 (10,5), refusé au-dessous de 7,04 (7), incertain de $[7,25 \text{ à } 10,25]$.

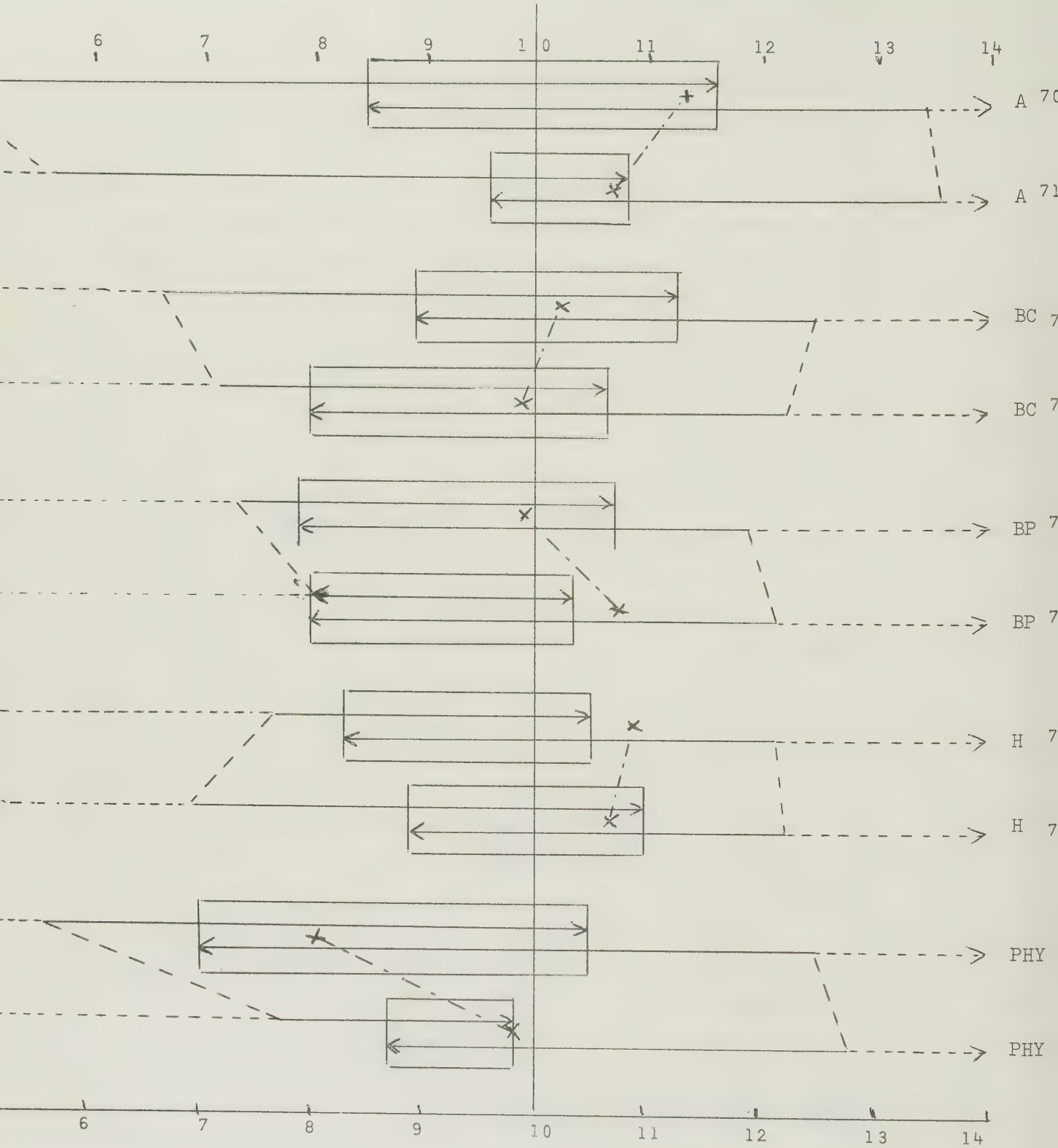
En 1971 : on était reçu au-dessus de 9,86 (10), refusé au-dessous de 8,72 (8,5), incertain de $[8,75 \text{ à } 9,75]$, la zone d'incertitude s'est rétrécie ! Alors qu'en BC une diminution de la variance attestait d'une augmentation de l'étendue de la zone d'incertitude, en PHY, une variation de même sens de la variance traduit une diminution de la zone d'incertitude (A et BP se conduisent identiquement).

Non seulement les variances ne sont pas comparables mais on ne peut même pas attribuer la même signification à un même sens de variation.

COMMENT EXPLIQUER CES DISCORDANCES ?

C'est un fait que chaque population peut se conduire différemment d'une discipline à l'autre. Ce fait se traduit par des différences de moyenne qui n'ont rien de scandaleux. Mais il est curieux de constater que les variances soient en désaccord dans leur signification. Quel est le résultat qu'on attend d'un examen, que les candidats se montrent bons ou mauvais ? Qu'il fasse la différence,

MOYENNES, ECARTS TYPES ET INCERTITUDE.



X = Moyenne de la discipline

_____ 2 écarts types des non reçus à partir de leur moyenne.

_____ 2 écarts types des reçus à partir de leur moyenne.

Les rectangles figurent les zones d'incertitude.

la discrimination, entre les acceptés et les refusés. Le "pouvoir de discrimination" d'une épreuve avait été jugé comme pouvant être évalué par la variance et la comparaison, sur ces données, entre BC et PHY aboutit à un échec. BC a été jugé comme discriminant moins bien dans l'ensemble, en 1971 qu'en 1970. La preuve avait été claire : au lieu de se demander que penser de la signification des notes $[9 - 11,25]$ on se posait la question pour l'intervalle $[8 - 10,75]$ ce qui n'était pas un progrès pour les refusés, alors que c'en était un pour les acceptés. Mais en BC, les moyennes n'avaient pas changé. En PHY, au contraire, la moyenne des refusés avait augmenté alors que celle des acceptés n'avait pas changé sensiblement. Les deux moyennes se sont rapprochées : la discrimination est donc moins bonne conformément à l'interprétation qui avait été initialement du sens de variation des variances. Mais ce rapprochement des deux moyennes enlève une partie de sa signification au rétrécissement de la zone d'incertitude. En PHY elle est passé des notes

$[7,25 - 10,25]$ en 1970, aux notes

$[8,75 - 9,75]$ en 1971.

Cet écart a diminué de deux points, la moyenne avait augmenté de deux points.

En A, de 1970 à 1971, la zone d'incertitude est passée

de $[8,75 - 11]$

à $[9,75 - 10,75]$.

Cet écart a diminué de 1,25 point, les moyennes se sont rapprochées de 1,02 point.

En BP, de 1970 à 1971, la zone d'incertitude est passée

de $[8 - 10,5]$

à $[8,25 - 10,25]$

Cet écart a diminué de 0,50 point, les moyennes se sont rapprochées de 0,41 point.

Dans ces trois disciplines, les variations des écarts entre les grandeurs des zones d'incertitude correspondent aux variations des écarts entre les moyennes des acceptés et des refusés. Que signifient alors les variations des variances des acceptés et des refusés ?

Ce que signifie toute variance et qu'exprime mieux sa racine : l'écart type. A savoir qu'elle indique la dispersion. Sur l'une quelconque des trois disciplines A, BP ou PHY en même temps que diminue l'écart entre les moyennes des acceptés et des refusés, les candidats se "resserrent" auprès de ces moyennes : les mauvais et les pires d'une part, les bons et les meilleurs d'autre part, ne sont pas

différenciés. Peut-être, en réalité, objectera-t-on, ces deux populations sont-elles homogènes ? Peut-être, en effet. Mais alors, plus les moyennes se rapprochent et plus les individus se groupent autour de ces moyennes, plus le "point" doit être significatif car, dans certaines régions de l'histogramme, une variation d'un point entraîne de considérables modifications des résultats. Or, la signification du point lui est donnée par la correction.

Finalement, la question se pose de la signification des variations de variance

1° - Ou bien l'examen est le résultat d'une mesure : quatre disciplines ont altéré leur technique puisque leur variance a diminué. Une l'améliore, mais alors, trois disciplines seulement ont altéré complètement leur technique puisque trois seulement diminuent la variance des acceptés et des refusés. Les deux autres, BC et H, ne mesurent pas de la même manière les acceptés et les refusés alors que ce doit justement être la mesure qui les discrimine.

2° - Ou bien l'examen n'est pas le résultat d'une mesure et une diminution de variance indique une population plus homogène : pour H, la population est moins homogène et ce sont les refusés qui le sont le moins. Pour BC, ce sont les acceptés. Pour A, BP et PHY la population est généralement plus homogène : voilà une différence dans l'interprétation. Quant à la grandeur de la variété, les valeurs des variances diffèrent toutes significativement les unes des autres aussi bien en 1970 qu'en 1971.

CONCLUSION

En admettant, faute de pouvoir le confirmer ou le contredire, que les moyennes soient justes et qu'il soit normal que les effectifs varient, que les épreuves soient en désaccord quant au nombre d'étudiants satisfaisants, il n'en reste pas moins vrai que l'écart entre les bons et les meilleurs, les mauvais et les pires, n'est pas jugé par tous de la même façon : le point n'a pas partout la même valeur.

Les grandeurs d'un examen ne sont pas des grandeurs mesurables par nature. Ce ne sont même pas des grandeurs comparables :

- faute d'un étalon défini: le candidat de référence
- faute d'une "unité d'intervalle" définie : le point.

CHAPITRE VI

UN JURY ET DES NOTES

Personne n'a jamais tout à fait tort:
même une horloge arrêtée a raison deux
fois par jour.

- DES REALITES I08
- COMMENT JUGE UN JURY II0
- EXAMEN D'UN JURY D'EXAMEN II2
- DES INFORMATIONS INCONNUES du Jury II7
- PROCES DU JURY II8
- APERCU DES RESULTATS (par classes) I2I
- CONCLUSIONS I29

DES REALITES

Chacune des épreuves étant passée, chacune étant notée, chaque département adresse au secrétariat des copies anonymes notées. L'anonymat levé, le secrétariat relève les notes (éventuellement affectées d'un coefficient) applique les règles concernant les épreuves partielles, et finalement, regroupe les résultats obtenus par chaque candidat dans chacune des disciplines et en fait le total. Cela représentait en 1970 (juillet) 26 pages de chiffres. En ne tenant compte que des seuls "candidats", c'est-à-dire des étudiants s'étant réellement présentés à chacune des épreuves, cela représentait 1 656 notes distribuées devant 276 noms : 1 380 notes d'épreuves et 276 notes totales. Chacune des 1 380 notes "élémentaires" étant, en réalité, déjà le total de plusieurs épreuves.

Au jour dit, un certain nombre de personnes se réunissent : chacune représentant un département. A chacune de ces personnes, le secrétariat remet une photocopie des pages du registre des résultats. Le registre lui-même, servant de pièce officielle est entre les mains du représentant de l'administration qui y portera les modifications et les sanctions prononcées par le jury. La dernière page sera finalement signée par chacun des membres de l'Assemblée ayant délibéré.

C'est au début de la séance que le jury prend connaissance de ce qu'il faut bien appeler des résultats quoique, en principe, c'est à lui-même que revient de prononcer les sanctions. Il faut bien

admettre l'impossibilité de prononcer une sanction (décision irrévocable) sans information : c'est le registre, constitué par l'administration (plutôt bien d'ailleurs) qui fournit ces informations. Quelles sont-elles ? Des notes. Si les informations sont fausses, les décisions prises ont peu de chance d'être justes. Ce n'est pas le secrétariat qu'on peut accuser de fournir de faux renseignements. L'étude attentive des résultats présentés ici a exigé de forts nombreux calculs échelonnés sur plusieurs mois : une seule erreur matérielle était relevée (0,25 point) et encore s'agissait-il d'un étudiant brillamment reçu ; une fausse note sur plus de 1 600 cela semble en deçà des marges tolérables.

COMMENT JUGE UN JURY

Cette année là, en juin, 128 candidats avaient totalisé au moins la moitié du maximum des points : ils ont été déclarés reçus. Ce seul fait indique que le jury fonctionne dans un seul sens : revenir sur des totaux trop bas. Il ne se préoccupe absolument pas des reçus qui le sont sur le seul total de leurs notes.

148 candidats n'avaient pas obtenu le fatidique "50 p. 100 du maximum des points". De ces 148, seuls 22 ont été examinés (à peu près 15 p. 100). Si le jury n'examine que des cas "non-reçus", il ne les examine pas tous.

La "moyenne" requise était de 55 points et n'ont été envisagés que les totaux au moins égaux à 53,75. A priori, il était donc admis que les notes étaient fiables à 98,87 p. 100 : la marge d'erreur admise était de 1,13 p.100. Exprimée sur 20, cette marge représente 0,226, soit, à peu près 1/4 de point. D'ailleurs, la moyenne (55) diminuée d'un quart de point par épreuve (5) donne bien 53,75. Ainsi le jury semble bien accorder une grande confiance aux notes dont il faut rappeler qu'elles constituent les seules informations disponibles.

Il est évident que, puisqu'il existe une règle du jeu, ce n'est pas au jury de la violer. Il est non moins évident que, recevoir un candidat ayant, ne serait-ce qu'un quart de point de moins que le minimum requis, constitue une transgression et la question est de savoir jusqu'où on peut aller. "Il faut bien une

limite" répète-t-on à l'envie. Certes, mais recevoir un candidat n'ayant pas le minimum requis, c'est dépasser les limites et il ne semble pas qu'il s'agisse là d'un abus de pouvoir. Dépasser pour dépasser, autant que ce soit raisonnable.
Raisonnable n'est pas, et n'a jamais été synonyme de timide, sauf crainte d'être déraisonnable. Et c'est être déraisonnable que de penser que la précision d'une telle notation dépasse 98 p. 100, aussi bien dans un sens que dans l'autre.

Cette année-là, il existait dans toutes les disciplines une note éliminatoire (5/20) qui signifiait que, quel que soit le total obtenu, le candidat devait réparer son échec à la session de septembre, les notes égales ou supérieures à la "moyenne" restant acquises. Ainsi, deux motifs pouvaient faire discuter le cas d'un candidat en jury : total insuffisant et note éliminatoire.

EXAMEN D'UN JURY D'EXAMEN

1° - Examinant des notes insuffisantes donnant des totaux insuffisants

Quatre cas seront analysés (la page est la référence du procès verbal).

Exemple	Page	A	BC	H	PHY/30	BP	Total	Décisions
1	13	10,75	9,5	11,25	16,5	6	54	+ BP + 1
2	19	14,75	13	11,25	9	6	54	+ H + 0,75 + BP + 0,25
3	19	10	9,5	9	16	10	54,5	+ BC + 0,5
4	26	15	7,5	10	10	11,5	54	-

Ces quatre exemples permettent de se faire une idée de ce que pense le jury de la précision des notes.

Première hypothèse raisonnable

Le quart de point n'est pas significatif. Les exemples deviendraient alors :

1er ex. : 11 + 9,5 + 11 + 16,5 + 6 = 54

2è ex. : 15 + 13 + 11 + 9 + 6 = 54

3è ex. : 10 + 9,5 + 9 + 16 + 10 = 54,5

4è ex. : 15 + 7,5 + 10 + 10 + 11,5 = 54

Cette hypothèse n'est pas confirmée par la décision du jury.

Deuxième hypothèse raisonnable

Le demi point n'est pas significatif (et a fortiori le quart de point !)

Les exemples deviendraient alors (soulignées les notes qui changeraient) :

1er ex. : 11 + 10 + 11 + 17 + 6 = 55
2è ex. : 15 + 13 + 11 + 9 + 6 = 54
3è ex. : 10 + 10 + 9 + 16 + 10 = 55
4è ex. : 15 + 8 + 10 + 10 + 12 = 55

Cette hypothèse n'est pas confirmée par la décision du jury.

Ce qu'on peut déduire des décisions du jury

du 1er exemple :

La marge d'erreur est de 1 point en BP
0,25 est significatif en A et en H
0,50 est significatif en BC et en PHY

du 2ème exemple :

0,25 est significatif en A
La marge d'erreur admise en H est 0,75
La marge d'erreur admise en BP est 0,25

du 3ème exemple :

La marge admise en BC est 0,5

du 4ème exemple :

La marge d'erreur admise en BC est inférieure à 0,5.

On peut objecter que le candidat n° 4 n'avait que 7,5 en BC, tandis que le n° 3 avait 9,5 et que le cadeau fait au n° 3 était attribué à un candidat "plus méritant". Mais alors le n° 1 avait, lui aussi, 9,5 ! BP admet qu'il existe une différence entre 11,5 et 12, sans quoi le n° 4 eût été vraisemblablement reçu mais BP estime par ailleurs que 6 = 7 (n° 1) ou 6 = 6,25 (n° 2). A l'évidence, les points ont été donnés pour "faire" 55. Sauf au n° 4 et la seule explication logique, c'est que 7,5 en BC, ce n'est vraiment pas assez ; (8 - 7,5) est différent de (10 - 9,5). C'est un résultat admissible : il signifierait que le point n'a pas la même valeur tout au long de l'échelle mais, alors, comment peut-on faire des additions ?

2° - Examinant un total insuffisant

Le rapprochement de deux paires de totaux insuffisants permettra de discerner le mode de jugement.

Exemple	Page	A	BC	H	PHY/30	BP	Total	Décisions
n° 4	26	15	7,5	10	10	11,5	54	-
n° 4bis	9	11,5	9	10	10,5	13	54	+ (BC + 1)
n° 5	14	11,25	13	11	10,5	8	53,75	+ PHY + 0,25 + BP + 1
n° 5bis	25	15	11	8,75	9	10	53,75	-

La paire (4-4bis) présente le double caractère d'un total égal et de défaillance dans les mêmes disciplines. Le n° 4 a déjà été examiné ; le n° 4bis a été reçu grâce à un point de plus en BC qui, selon les cas, accorde ou non une signification au demi point (n° 1 et n° 3) et, dans le cas présent estime que 9 = 10 !

Le candidat n° 5 a été reçu grâce à BP qui estime ici, qu'il n'y a pas de différence significative entre 8 et 9, pas plus qu'entre le 6 et le 7 dans l'exemple n° 1 mais contrairement à son jugement dans l'exemple n° 2 ! Dans ce cas, PHY a estimé que 10,5 pouvait se lire 10,75, mais, en 5bis, 9 = 9 pour PHY et 8,75 = 8,75 pour H qui n'hésitait pourtant pas (ex. 2) à transformer un 11,25 en 12.

Jusqu'ici, BC, H et BP ont été pris en flagrant délit de contradiction sur la valeur de leurs propres notes car des décisions n'avaient pas été prises en jury sans l'assentiment du représentant du département.

3° - Examinant une note éliminatoire

C'est PHY qui en a fourni le plus : 88. Cette discipline étant notée sur 30, l'éliminatoire est inférieure à 9.

Quatre candidats se sont trouvés en délibération avec 8,5 en PHY et un total général suffisant (supérieur à 55). Deux fois ces notes insuffisantes en PHY étaient accompagnées d'une autre note inférieure à la moyenne : ces deux candidats, doublement insuffisants, dont une fois gravement, ont été reçus. Deux fois, la note insuffisante en PHY était isolée : l'un des deux candidats a été reçu, l'autre non !

Exemple	Page	A	BC	H	PHY/30	BP	Total	Déci-sion
6	9	13,25	11	15,5	8,5	11,5	59,75	-
7	14	13	10,5	13,5	8,5	10	55,5	+

Il est extrêmement difficile de discerner en quoi le demi point qui sépare de la note fatidique le numéro 6 est différent de celui du numéro 7. D'autant, qu'à se fier au total, le candidat refusé serait plutôt meilleur... A moins que le total ne présente pas d'intérêt ! Et pourtant, il doit en avoir un puisqu'un candidat (page 21) n'ayant eu que 8 en PHY (et 7/20 en BP) a été reçu : son total était 60. Peut-être que ce quart de point de plus au total compensait le demi point de moins en PHY et la mauvaise note en BP, sans quoi on ne voit pas en quoi ce candidat reçu est meilleur que le numéro 6. D'ailleurs, le total doit compter puisque (page 1) un candidat a été reçu avec 7 sur 30 en PHY mais un total de 68,5 tandis que (page 16) un autre été refusé avec 7 sur 30 en PHY mais un total de 54,5. Il faut ajouter que (page 22) 6 sur 30 était jugé acceptable avec un total de 56,5 !

Evidemment, l'un des deux candidats refusés avec 7 sur 30 en PHY a un total insuffisant : il n'a que 54,5 mais il faut voir comment :

(A) 13,5 + (BC) 11,5 + (H) 12,5 + (BP) 10 + (PHY) 7 = 54,5

Or, si on se réfère à quelques uns des exemples précédents (mais pas n'importe lesquels) 0,5 n'est significatif ni en BC, ni en H et le total aurait aussi bien pu être 55,5.

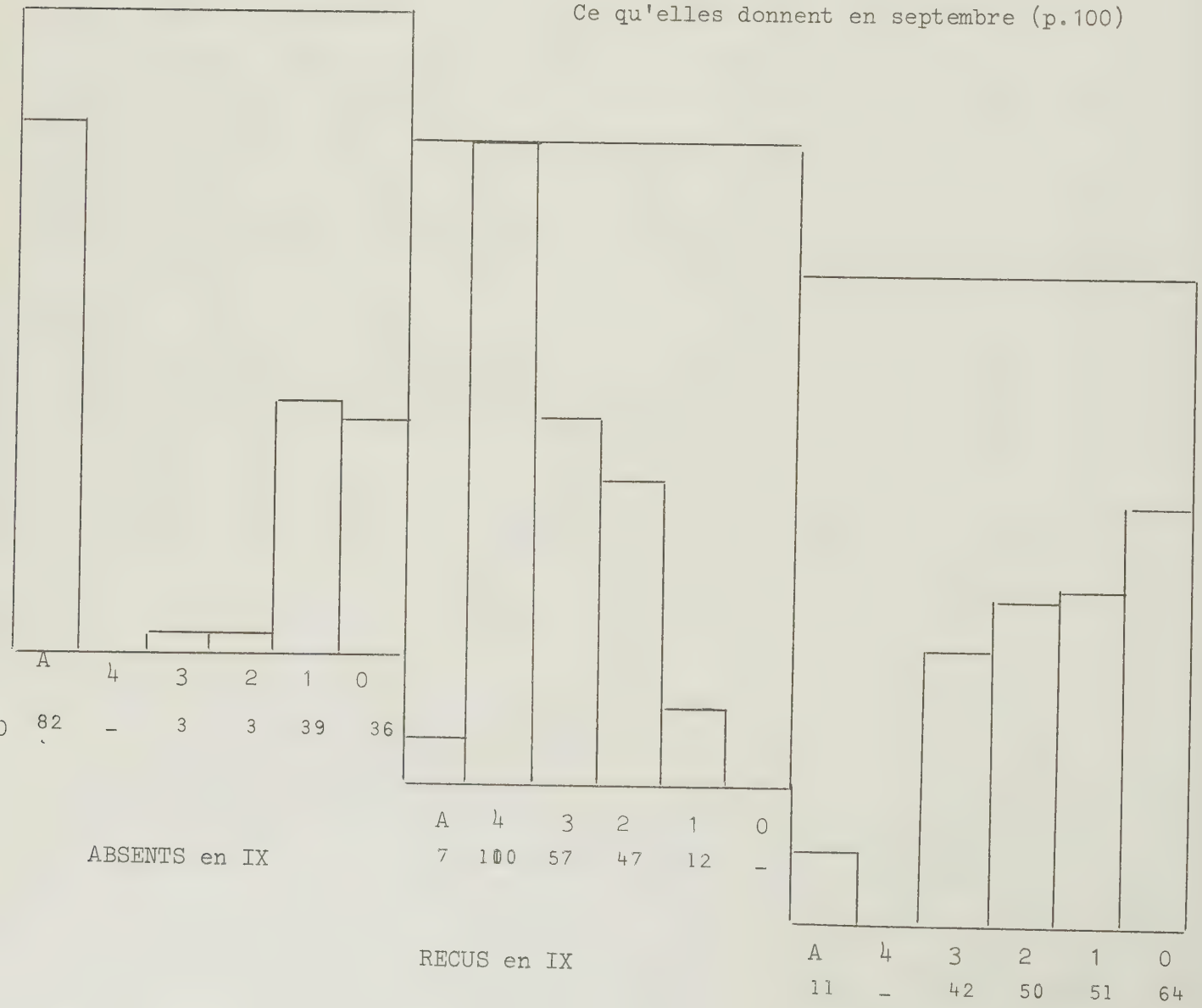
Il semble donc bien que, pour le jury lui-même, la signification du point ne soit pas très claire et que la logique ne préside pas aux débats.

LES CLASSES DE JUIN 70

(fréquences absolues)

Abs:56 $_5$C 54 $_4$C 59 $_3$C 71 $_2$C 33 $_1$C 34 $_0$C 25

Ce qu'elles donnent en septembre (p.100)



DES INFORMATIONS INCONNUES DU JURY

Pour ne retenir que des informations fournies par les résultats même de l'examen, le jury n'avait sous les yeux que 1 656 notes (sans compter celles des candidats n'ayant pris part qu'à une partie des épreuves) groupées par ordre alphabétique et il ne disposait que d'un temps trop limité pour lire, ce qui s'appelle lire, les résultats. Tout au plus savait-il que 128 candidats seraient certainement déclarés "reçus" puisqu'ils avaient un total suffisant sans note éliminatoire (l'éliminatoire a été supprimée depuis).

Il ne savait pas que 17 p. 100 des étudiants ne s'étaient pas présentés ce qui représente le troisième rang d'importance des sept classes possibles. Sans doute avait-il la notion d'un classement possible des candidats sur d'autres critères que le total. Mais il ne savait rien du classement réel. Faute d'en avoir une représentation synthétique, il ne savait même pas comment il avait délibéré : c'est pour cette raison que la page du procès verbal a toujours été notée dans les exemples précédemment cités ; à l'évidence, le cas des deux 7 sur 30 en PHY ayant abouti à des décisions contradictoires peut s'expliquer par l'écart des candidats : l'un en page 1, l'autre en page 16, ce qui représente un intervalle de temps notable. Il est hautement probable que le simple rapprochement de ces deux cas eût interdit soit la réception, soit l'élimination de l'un des deux.

Enfin, le jury ne délibérant, en fait, que sur le total, ignorait à quel point ses décisions de juin influeraient sur les résultats de septembre et à quoi, exactement, ressemblaient ses décisions.

PROCES DU JURY

Il est très facile de se moquer du jury en soulignant ses contradictions. En prenant les débats par le même bout que lui, il est possible d'expliquer son attitude.

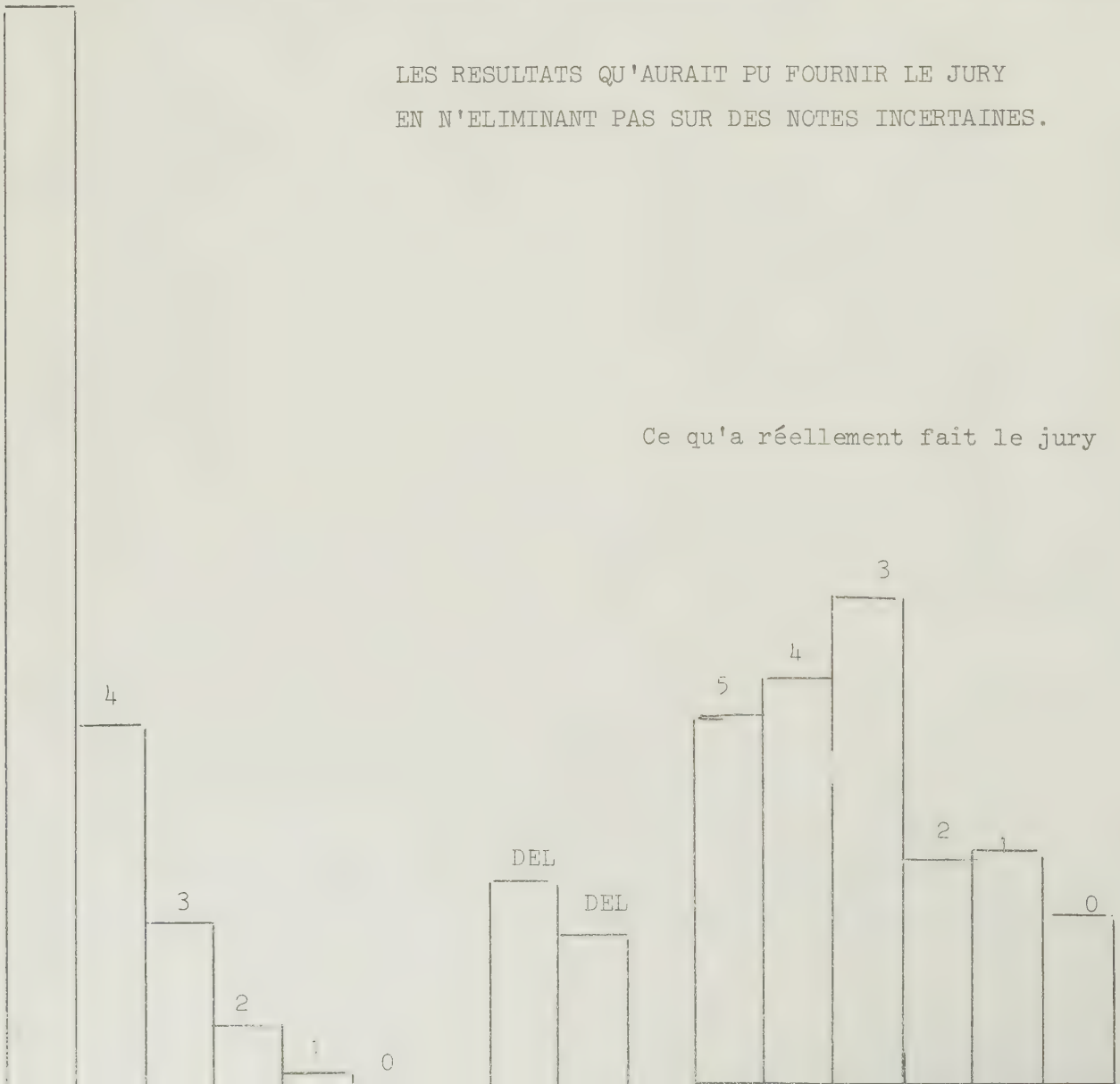
D'abord le jury, est, en fait, une assemblée de comptables ayant une mentalité d'avocats. Tous les candidats dont il a à connaitre ont, au moins, quelques pécadilles sur le dossier. Le rôle, assez chevaleresque, de redresseur de torts entretient une certaine bonne conscience. Tels candidats sont-ils injustement reçus ? Je n'y peux vraiment rien : ils avaient de bonnes notes. Tels autres ont-ils été refusés dans des conditions discutables ? Que voulez-vous que j'y fasse : c'est encore beau d'avoir reçu tel ou tel avec les notes qu'ils avaient ! N'y aurait-il qu'un seul reçu après délibérations que le jury serait justifiable : ceux qui ont été reçus sans délibération avaient de bonnes notes et tous ceux qui ont été refusés l'ont été avec des notes insuffisantes ; que celui qui a été reçu s'estime heureux ! A la limite, on pourrait suggérer que tout acte supplémentaire à l'apposition de la signature finale est vertueux.

Ensuite, cette assemblée est mal informée. Enfin, elle est, d'une certaine façon, juge et partie. En effet, les membres qui la constituent ont été des juges primaires, ils ont constitué ou participé à la constitution du dossier d'une épreuve : ils ont été correcteurs, plus ou moins responsables de la technique de correction, et, en règle, plutôt largement. Ce qui est demandé à ces gens, c'est de reconnaître que leurs notes sont discutables, c'est admettre que

Possibles / els	CL.5	CL.4	CL.3	CL.2	CL.1	CL.0	Douteux	Totaux
CL.0		5	5	7	2		6	25
CL.1	3	9	10	2			10	34
CL.2	11	9	7				6	33
CL.3	43	19	2				7	71
CL.4	47	11					1	59
CL.5	54							54
Totaux	158	53	24	9	2		30	276

LES RESULTATS QU'AURAIT PU FOURNIR LE JURY
EN N'ELIMINANT PAS SUR DES NOTES INCERTAINES.

Ce qu'a réellement fait le jury



certaines d'entre elles ne veulent rien dire. Il leur est demandé de se déjuger. Il n'est pas du tout certain qu'ils n'en soient pas capables, mais entre le premier jugement et le second, la seule information supplémentaire dont ils disposent c'est la grandeur de l'effectif des reçus, grandeur qui dépend justement de leur premier jugement. Hormis les cas sociaux, on ne voit vraiment pas ce qu'ils pourraient sérieusement discuter. Il semble bien, en vérité, que les contradictions, précédemment soulignées avec quelque complaisance, soient l'indice d'un certain désarroi. Si le jury est condamnable, quelques circonstances atténuantes peuvent être discutées et, si on le condamne finalement, ce ne sera pas sur ce qu'il fait mais sur la manière dont il le fait, dans la mesure où il serait démontré qu'il peut mieux faire.

APERCU DES RESULTATS

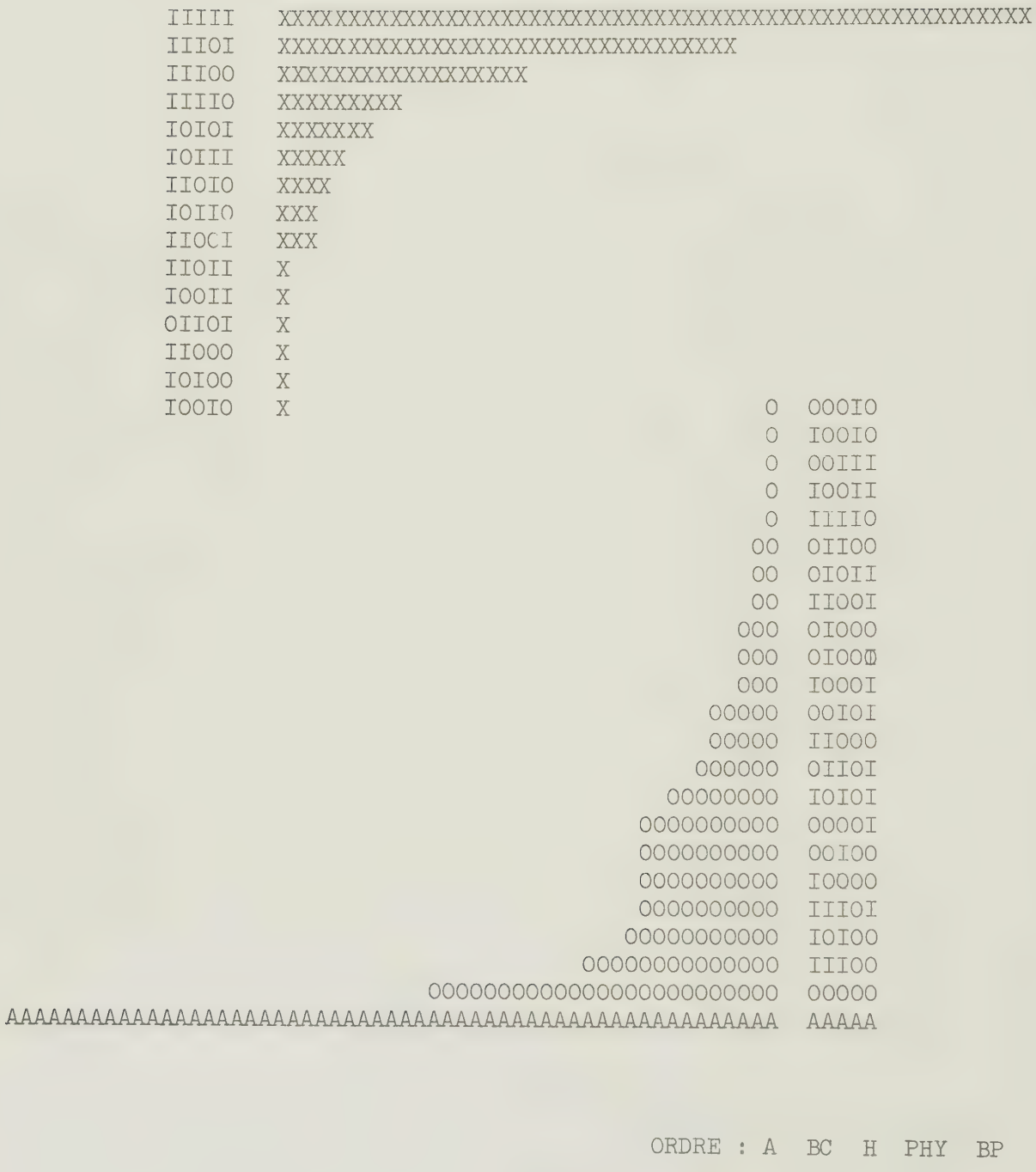
Deux sortes d'étudiants sont à considérer : ceux qui se présentent à toutes les épreuves et qui sont appelés candidats, et les autres, qui ne se présentent pas, ne serait-ce qu'à une épreuve : ce sont les absents.

Les candidats peuvent être "acceptés" dans un nombre quelconque d'épreuves, allant, ici, de zéro à cinq : ainsi se trouvent définies six classes. La classe 5C représente la réussite à chacune des cinq épreuves et se représente par "11111" tandis que la classe oC, échec dans chacune des cinq épreuves, se représente par "ooooo". Hormis ces deux classes extrêmes d'accords parfaits, chacune se subdivise en "catégories" caractérisées par la nature de l'échec ou des échecs. La classe 4 est caractérisée par un échec et subdivisée en cinq catégories selon que l'échec (0) est placé sous telle ou telle rubrique. Il y a dix catégories en 3C et en 2C, cinq en 1C. En tout, 32 (soit 2^5) dont deux seulement sont univoques : ${}_5C$, réussite partout ; oC : échecs partout. Dans chacune des quatre autres il y a des combinaisons diverses d'échecs et de réussites. A priori, dans chacune des 30 catégories (32-2) où se combinent bons et mauvais résultats, deux sanctions peuvent être portées : reçu ou refusé. Au total, 62 catégories sont finalement possibles (26 - 2).

Qu'il soit bien clair qu'à partir de la position où on accepte de recevoir un candidat ayant "écrit son message" autrement que "11111", c'est le total qui tranche et il devient parfaitement acceptable de recevoir un candidat "10000" ou l'une quelconque des combinaisons représentée par 10000, ce qui s'est d'ailleurs produit. Cela signifie qu'une seule réussite peut compenser quatre échecs si on admet qu'une note inférieure à 10 sur 20 est un indice d'échec. Par conséquent, dans l'hypothèse d'un examen à cinq épreuves notées

LES 18 FACONS D'ETRE RECU ET LES 22 + 1 FACON D'ETRE REFUSE

EN JUIN 1970.


IIIII XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
IIIOI XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
IIIOO XXXXXXXXXXXXXXXXX
IIIIO XXXXXXXXX
IOIOI XXXXXXX
IOIII XXXXX
IIOIO XXXX
IOIIO XXX
IIOOI XXX
IIOII X
IOOII X
OIIOI X
IIOOO X
IOIOO X
IOOIO X
O OOOIO
O IOOIO
O OOIII
O IOOII
O IIIIO
OO OIIOO
OO OIOII
OO IIOOI
OOO OIOOO
OOO OIOOO
OOO IOOOI
OOOOO OOIOI
OOOOO IIOOO
OOOOOO OIIOI
OOOOOOOO IOIOI
OOOOOOOOOO OOOOI
OOOOOOOOOO OOIOO
OOOOOOOOOO IOOOO
OOOOOOOOOO IIIOI
OOOOOOOOOOO IOIOO
OOOOOOOOOOOOOO IIIOO
OOOOOOOOOOOOOOOOOOOOOOO OOOOO
AAAA

sur 20

on recevra : 14 + 9 + 9 + 9 + 9

ou : 9 + 14 + 9 + 9 + 9

mais pas : 9 + 9 + 9 + 9 + 9

Le 9 sur 20 dans la première discipline, par exemple, sera jugé sans importance si la seconde a été cotée 14 et rédhibitoire si la seconde a été cotée 9. Si on considère la cinquième discipline, le candidat ayant obtenu 14 dans la première ou la seconde sera déclaré "bon". Le tableau précédent montre donc des "bons" et des "mauvais" 9. Cet abord de la question prend toute son importance lorsqu'on a considéré avec un peu d'attention chacune des classes.

Classe 5C

Avec 54 candidats, elle représente 20 p. 100 de l'effectif. C'est, par excellence, la "bonne classe" : toutes les disciplines sont en accord, les candidats ont été satisfaisants partout. On y trouve :

26 p. 100 des reçus en Anatomie

31 p. 100 des reçus en Biochimie

27 p. 100 des reçus en Histologie

64 p. 100 des reçus en Physiologie

35 p. 100 des reçus en Biophysique

La moyenne de ces candidats est élevée par comparaison avec la moyenne vraie de l'épreuve ($\mu.\bar{X}$)

	μ 5C	μ $\bar{X}$
A	16	11
BC	13	11
H	14	11
PHY	13	8
BP	13	10

On peut remarquer :

- la moyenne relativement très élevée en A et en PHY
- la moyenne relativement modeste en BC
- une anomalie, la modestie de l'effectif. Seulement 20 p. 100 des candidats sont déclarés unanimement satisfaisants. Cette classe occupe le quatrième rang d'importance numérique.

ANALYSE DES CLASSES

	MOYENNES				VARIANCES			
	5^C	TOTAL	1_4^C	$_oC$	5^C	TOTAL	1_4^C	$_oC$
A	15.69	11.4	11.22	3.82	13.30	21.04	15.59	4.31
BC	13.38	10.56	10.34	6.16	4.39	11.91	23.14	2.39
BP	12.87	10.03	9.66	6.84	6.27	8.91	6.53	4.99
H	13.61	11.00	10.75	7.35	4.00	6.96	4.83	1.72
PHY	13.43	7.61	6.50	3.77	11.35	21.78	14.34	5.65

CORRELATIONS (r)

	5^C	TOTAL
A RESTE	0.586	0.707
BC RESTE	0.423	0.554
BP RESTE	0.067	0.489
H RESTE	0.558	0.903
PHY RESTE	0.323	0.625
A BC	0.561	0.601
A BP	0.071	0.385
A H	0.594	0.667
A PHY	0.381	0.571
BC BP	-0.048	0.284
BC H	0.548	0.645
BC PHY	0.229	0.461
BP H	0.069	0.380
BP PHY	0.048	0.437
H PHY	0.292	0.503

Effectif TOTAL = 276
Effectif $_4^C$ = 197
Effectif $_5^C$ = 54
Effectif $_oC$ = 25

Ce sont les notes vraies qui ont servi et non celles fourni par les histogrammes.

Classe 4C

Avec 59 candidats, elle représente 21 p. 100 de l'effectif (3ème rang). Sur les cinq combinaisons possibles, quatre sont effectivement réalisées. Personne n'a été déclaré "satisfaisant partout sauf en Anatomie". Cela permet de se faire une idée des notes suffisantes dans quatre disciplines. Ce sont :

- en Biochimie ..... 8,5 sur 20
- en Histologie .... 8,5 sur 20
- en Physiologie ... 4 sur 20 (6 sur 30)
- en Biophysique ... 6 sur 20

Il est intéressant de rapprocher ces notes minimales admises par le jury de celles qui avaient été prévues : (cf. chapitre II)

	minimum réel	minimum prévu
BC	8,5 = 9	8,94 = 9
H	8,5 = 9	8,34 = 8,5
PHY	4/20 = 6/30	7,04 = 7/20
BP	6	7,79 = 8

On peut voir qu'en Biochimie et en Histologie les prévisions sont très proches de la réalité. En Physiologie et en Biophysique, en revanche, le minimum réel est sensiblement inférieur à celui qui avait été prévu. Cela revient à dire que, en ce qui concerne ces deux disciplines, le jury s'est montré beaucoup plus indulgent que ce qui semblait admissible. Mais, ce qui semblait admissible, le jury l'ignorait. L'absence de citation de l'Anatomie s'explique très bien si on veut bien se reporter à l'histogramme correspondant (I). Il est encore intéressant de souligner que 81 p. 100 des candidats de cette classe ont été déclarés reçus. Cela signifie que des notes, parfois très basses, ont été estimées comme ne devant pas mettre obstacle à la progression du candidat d'un grade à l'autre. L'examen des effectifs cumulés des acceptés en 5C et en 4C, discipline par discipline, n'est pas sans intérêt puisque c'est à ce niveau que se "creusent" le plus les écarts. Le tableau suivant représente, discipline par discipline, le pourcentage des reçus en 5C, 4C, et en $^{5}_{4}C$ (ou 5C + 4C).

	$_5C$	$_4C$	$\frac{5}{4}C$
A	26	29	55
BC	31	31	62
H	27	29	56
PHY	64	19	83
BP	35	31	66
reçus	38	34	72

Un dernier renseignement peut encore être fourni par cette classe. Sur onze candidats refusés, 10 l'ont été pour la seule physiologie, puisque, par définition, ils étaient acceptés partout ailleurs. Or, d'une part, la moyenne de ces refusés, en Anatomie est plus basse que celle des 48 reçus de la même classe, et, d'autre part, la moyenne en Physiologie est aussi basse que celle des candidats appartenant à la classe oC. Il est aussi difficile de tenir le jury pour responsable de ce fait que d'absoudre totalement le système d'évaluation.

Classe 3C

Elle occupe le premier rang d'importance numérique (26 p. 100 des candidats). Le fait a déjà été signalé et il pose avec une acuité particulière la question de la justesse des corrections puisque, à en croire les notes, les candidats joueraient, le plus souvent, sur deux impasses. Par ailleurs, il semble scandaleux que 51 p. 100 de ces candidats déclarés doublement insuffisants, aient été finalement déclarés reçus.

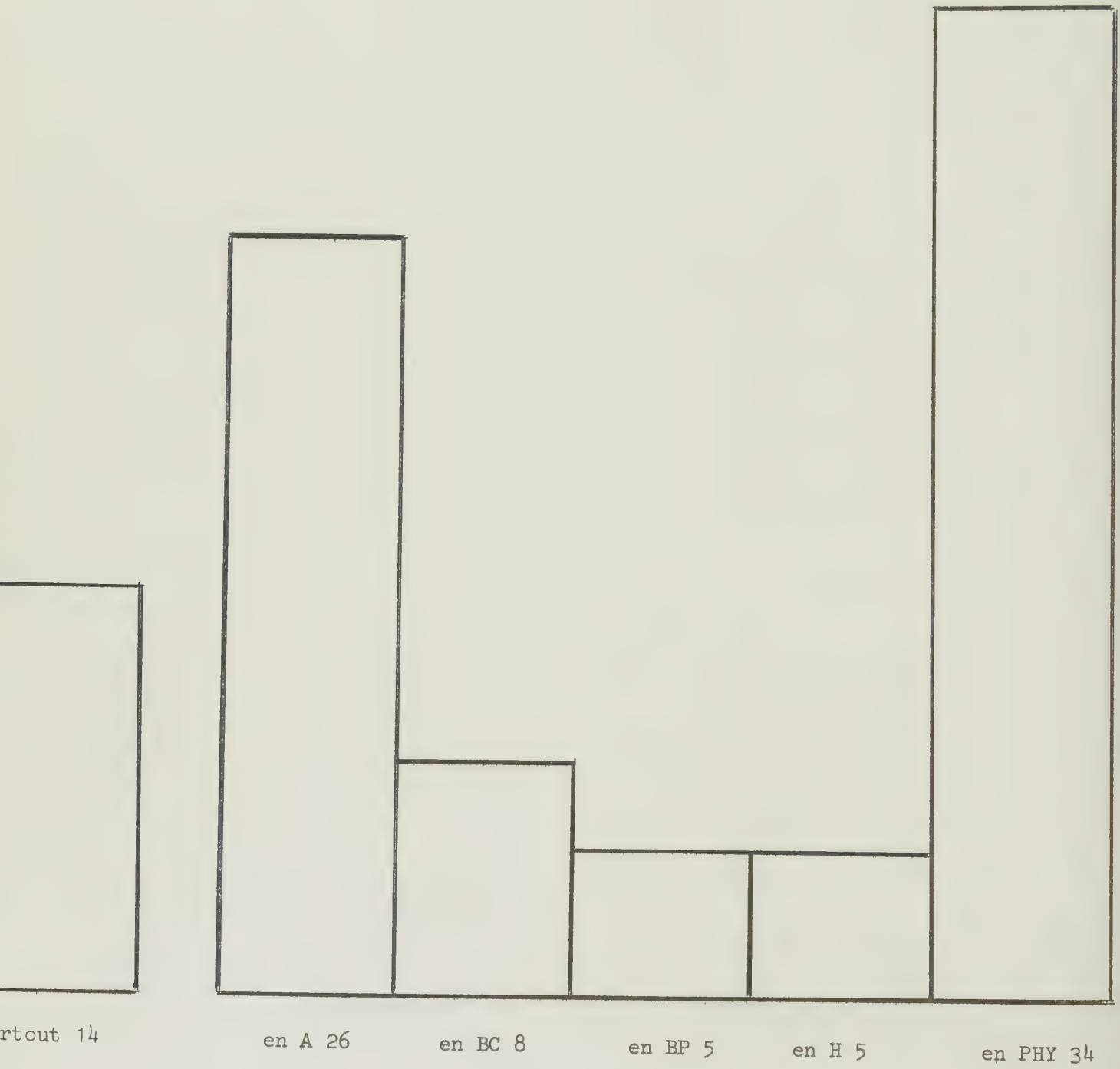
Plus encore, cette classe semble une caricature de tout ce qu'on peut reprocher à l'examen : incertitude, indécision, imprécision et incohérence.

1° - En Physiologie, la moyenne des reçus est inférieure à celle des refusés !

2° - En Anatomie, en Biochimie et en Histologie, la moyenne est inférieure à la moyenne générale : ces candidats sont relativement mauvais. Mais la valeur de cette "mauvaise moyenne" vaut d'être

ANALYSE DES ABSENTS

(Discipline par discipline,compte non tenu des absences multiples.)



A et PHY ont un point commun ce sont des disciplines nouvelles.

soulignée : c'est 10 sur 20 !

3° - Le jury offre un festival de contradictions :

A	reçoit à 6,75	et refuse à	8,5
BC	reçoit à 7,5	et refuse à	8,5
H	reçoit à 8,25	et refuse à	8,75
PHY	reçoit à 9/30	et refuse à	13/30
BP	reçoit à 6	et refuse à	9,5

Cette classe est un monument d'ambiguité.... et c'est la plus nombreuse.

CONCLUSIONS

Le jury substitue à la décision partielle d'un correcteur une décision collective qui s'impose alors souverainement à tous. La décision prise par un jury est un jugement porté à partir d'informations données.

Les informations dont dispose le jury sont les décisions partielles des correcteurs et ne valent que ce que valent les corrections ; elles s'expriment sous formes de notes qui n'ont, pour personne semble-t-il, une signification claire. Dans ces conditions, il semble difficile d'admettre que la décision du jury soit claire.

1) Les notes sont INDECISES et deux arguments appuient cette proposition. D'une part, et dans certaines disciplines, le nombre de "10/20" le suggère si on comprend de telles notes comme exprimant "juste suffisant" et signifiant "ni accepté ni refusé". D'autre part, le nombre de notes changées tendrait à le prouver : telle valeur n'a pas une signification arrêtée puisque son changement se fait, en quelque sorte, à la demande.

2) Les notes sont IMPRECISES. Le fait de les changer, sans nouvelle correction, montre qu'il n'est pas attaché une signification donnée à une valeur donnée. D'autre part, les membres du jury qui sont des correcteurs attachent manifestement une importance différente à deux notes s'exprimant par le même chiffre.

3) Les notes sont INCERTAINES. Celà est évident si on admet qu'elles sont indécises et imprécises. L'examen

qui a presque toujours servi de référence (juin 70) était caractérisé par l'existence d'une note éliminatoire qui, logiquement, signifie "candidat absolument insuffisant" or, un certain nombre de candidats ont été reçus malgré une note éliminatoire. C'est encore un fait que d'avoir constaté que la zone d'incertitude réelle était encore plus large que ce qui avait été prévu.

Les notes sont INCOHERENTES. La plus belle illustration qu'on en puisse donner, c'est que des candidats ont été reçus avec un brevet d'insuffisance. Mais l'incohérence majeure consiste à faire une addition donnant un résultat certain à moins de 2 % selon la formule :

INDECISION + IMPRECISION + INCERTITUDE
= DECISION CERTAINE A 98,87 p. 100.

Au total, le verdict du jury peut être considéré comme un préjugé aggravant les fautes des jugements élémentaires puisqu'il ne juge pas des notes mais à partir des notes et que son seul souci, finalement, c'est de modifier une note totale incohérente sans modifier en quoi que ce soit les causes d'une telle incohérence.

CHAPITRE VII

LE MAL

" Nous devons avoir plus de vénération
pour les vérités évidentes que
d'obstination pour les idées reçues."

PASCAL.

BILAN 132

Les questions qui se posent, comment y répondre, où en sommes nous?

UN NOMBRE : La note 134

DEUX MYTHES 137

La moyenne (137), la courbe de Gauss (138), les deux ensemble (140).

VANITE DE LA STATISTIQUE 141

LA NOTE N'EST PAS TOUT 145

CONCLUSIONS 146

ERRATUM : Page 145, ligne 8, lire : "...l'intégrale d'une FONCTION complexe...".

BILAN

LES QUESTIONS QUI SE POSENT

1° - Tel candidat est-il refusé ou accepté ?

2° - Qu'il soit l'un ou l'autre, est-il "bon" est-il "excellent" ? "mauvais" ou "très mauvais" ?

COMMENT Y REPONDRE

1° - Tel candidat est-il, oui ou non, conforme à la référence du candidat accepté ? Ce qui suppose cette référence définie.

2° - De combien s'éloigne-t-il de la référence, dans un sens ou dans l'autre ? Ce qui suppose défini un degré d'éloignement de la référence.

OU EN SOMMES NOUS ?

Dans l'ignorance la plus complète : ces définitions sont ignorées et cette ignorance même est ignorée.

Au début du XVII° siècle, lorsqu'un physicien parlait de "zéro degré" (référence des températures), il fallait l'interrompre pour s'enquérir d'où il venait. Etait-il Florentin ? Pour le comprendre il fallait entreprendre le voyage et descendre dans la fameuse cave. Etait-ce FAHRENHEIT ou l'un de ses disciples ? Il fallait s'enquérir de la composition du mélange. Etait-il Britannique ? Il fallait savoir ce que représentait cet éloignement de 65° de la température de congélation de l'eau. Pour les Florentins et les Prussiens, zéro, cela voulait dire qu'il faisait froid. Mais tant qu'on n'avait pas trempé son doigt dans le mélange ou exposé sa personne dans la cave, on ne se rendait pas compte à quel point. Pour les Anglais, zéro cela voulait dire qu'il

faisait très chaud. Pour REAUMUR et, plus tard, CELSIUS, zéro, cela pouvait se définir correctement : d'une telle façon que tout le monde comprenne. La "moyenne" pour nos modernes examinateurs cela se sent mais ne se définit pas. Pire, cela s'exprime par un nombre (10/20) mais sans définition.

Le terme "degré" est assez vague en dehors de son sens propre de marche d'escalier ou de barreau d'échelle. Vague mais significatif : quand on est sur le onzième degré, on est plus haut que si on était sur le dixième et plus bas que sur le douzième. Ceci, à condition de convenir que l'ordinal le plus bas soit aussi le plus petit. La différence de hauteur entre deux degrés consécutifs n'est pas toujours définie et cela n'a pas nécessairement d'importance. On sait qu'en alpinisme, une difficulté de 5ème degré se surmonte moins facilement qu'une difficulté du 4ème, qu'un vent 8° BEAUFORT est plus "fort" qu'un vent 7°, un séisme de 10° de l'échelle RICHTER est plus "violent" qu'un séisme de 9° .... etc. L'expérience montre, depuis longtemps et dans les domaines les plus divers, qu'une copie notée "11" n'est pas nécessairement meilleure qu'une copie cotée 10, voire 9 ou 8. Elle est peut-être meilleure qu'une autre cotée 15.

UN NOMBRE

Le nombre par lequel le "correcteur" exprime son jugement est une note. Dans la réalité, une épreuve étant constituée de plusieurs questions et chacune d'elles étant notée indépendemment par rapport à un maximum qui n'est pas forcément le même de l'une à l'autre, la note d'une épreuve est une addition dont la somme peut être affectée d'un coefficient aux fins de normaliser les expressions finales ramenées au même maximum. Dans une question, il n'est pas habituel de déterminer à l'avance ce qu'on attend et qui représenterait le minimum exigé pour déclarer le candidat "juste suffisant". Dans la réalité, c'est la lecture de plusieurs réponses qui donne "une idée" de la réponse moyenne. C'est par rapport à ce "portrait robot de réponse" que sera mise la note : elle ne s'attribue pas, elle s'invente.

Le point important, en ce qui la concerne, est sa fiabilité. Soit une épreuve, cotée de zéro à vingt, et dix, la "moyenne". La notation par points entiers, non fractionnés, amène à déclarer "reçu" celui qui a dix ou onze et non reçu celui qui n'obtient que neuf. Si la notation couvre réellement l'intervalle zéro-vingt, un point représente un vingtième de la grandeur totale, soit 5 p. 100. En ne discutant pas le critère de la moyenne, éliminer un candidat à neuf sur vingt, c'est être certain d'une précision de notation d'au moins 95 p. 100 : ce qui semble ambitieux pour une telle "mesure". Il en est évidemment de même pour un candidat reçu à onze sur vingt. En réalité, la précision doit être encore plus grande car le "neuf" pouvait être un "dix limite". Par exemple, "9.5 plus ou moins epsilon" arrondi à neuf ou à dix. La précision doit être alors de 97.5 p. 100 et elle doit être de 98.75 p. 100 si on note au demi point et de 99.375 p. 100 si on note au quart de point.

Inversement, exiger 90 p. 100 de précision, ce qui est beau

	−	+	○		−	+	△
MAXIMUM 5	≤ 2	≥ 3	2.5 (1)		≤0.40	≥0 60	0.200
10	≤ 4	≥ 6	5		≤0.40	≥0.60	0.200
20	≤ 7	≥13	8	12	≤0.35	≥0.64	0.300
25	≤ 9	≥16	10	15	≤0.36	≥0.64	0.280
30	≤11	≥19	12	18	≤0.366	≥0.633	0.267
40	≤15	≥25	16	24	≤0.375	≥0.625	0.250
50	≤19	≥31	20	30	≤0.380	≥0.620	0.240
80	≤31	≥49	32	48	≤0.3875	≥0.6125	0.225
100	≤39	≥61	40	60	≤0.39	≥0.6	0.220

(1) Cette note est impossible puisqu'il ne s'agit pas d'un nombre entier.

et peut-être utopique avec une telle méthode, c'est déclarer :

. Reçu le candidat qui a plus de 12/20

. Non reçu celui qui a moins de 8/20

. Incertain celui qui a 8, 9, 10 ou 11.

En admettant cette incertitude (optimiste) de 10 p. 100, un candidat serait classé reçu (+), non reçu (-) ou douteux (0) d'une manière différente selon le nombre maximum de points attribués à l'épreuve ainsi que le montre le tableau.:

. Dans la première partie (gauche) sont reportées en trois colonnes, les notes limites :

. supérieure des non reçus (-)

. inférieure des reçus (+)

. extrêmes des incertains (0)

. Dans la seconde partie, l'expression des résultats comparés à UN : le delta représente un indice d'incertitude.

. Neuf rangées représentent neuf valeurs choisies du maximum.

. La notation se fait impérativement par points entiers.

On peut remarquer que les deux valeurs du maximum les moins entachées d'erreurs sont cinq et dix et que la plus grande incertitude affecte la notation sur vingt qui se trouve être généralement choisie.

Entre cinq et dix, on peut choisir "cinq" en estimant souhaitable d'éliminer tout doute ou "dix" en estimant souhaitable de conserver un doute. Pour tout autre maximum choisi, il faut admettre la marge minimale d'incertitude impliquée par le choix. Ne pas en tenir compte rend douteux les résultats moyens qui sont les plus nombreux.

DEUX MYTHES

1) Le mythe de la "moyenne"

Voilà un terme d'une ambiguïté quasi parfaite et qui, malgré celà, régit les habitudes françaises en matière d'examen. On peut lui attribuer au moins quatre sens :

a/ valeur moyenne de la population (moyenne vraie)
b/ valeur minimale exigée pour être déclaré suffisant (seuil de suffisance)
c/ valeur située à égale distance de la nullité et de la perfection (médiane du correcteur)
d/ valeur située à égale distance du meilleur et du pire observés (médiane de la population).

La moyenne vraie est le résultat d'une opération qui consiste d'abord à prendre connaissance de la valeur de chaque individu, d'en faire la somme ensuite et, enfin, de diviser cette somme par le nombre des candidats. Cette valeur moyenne est imprévisible par définition. Or les calculs de vingt moyennes d'épreuve ont montré que dix huit fois leurs différences n'étaient pas significatives. Il est difficile de croire que ce genre de résultats soit le fait du seul hasard.

La valeur minimale exigée pour être déclaré suffisant est constamment "la moitié du nombre des points pouvant être attribués au maximum" (10/20) ce qui se traduit souvent par la moitié des connaissances exigées. Une copie "moyenne" est une copie à moitié bonne, un candidat "moyen" est celui qui a satisfait à la moitié des exigences. A la moitié, il est "satisfaisant". Il est satisfaisant de faire les choses à moitié. On est suffisant lorsqu'on est à moitié parfait.

La valeur située à égale distance de la nullité (zéro) et

de la perfection (20/20) c'est, en fait, la médiane du correcteur : c'est le point milieu qui sépare en deux parties égales l'échelle des notes qu'il est possible d'attribuer ; c'est la "médiane du possible".

La valeur située à égale distance du meilleur et du pire observés, c'est le point milieu qui sépare en deux parties égales l'intervalle entre la plus basse note réellement attribuée (6/20 par ex.) et la plus haute note réellement attribuée (16/20 par ex.). C'est la "médiane du réel".

Il est constant que soient confondus "seuil de suffisance" et "médiane du possible" : ainsi la note (marque) détermine le résultat au lieu de l'exprimer. C'est une première pétition de principes.

Il est habituel que soient confondus la médiane du possible et la moyenne vraie et ceci aussi bien si le correcteur "estime" la population "bonne" ou "mauvaise". Ainsi la moyenne apparait-elle comme une valeur pré établie et non comme l'expression d'une grandeur "libre". Ce qu'il s'agit de montrer est tenu pour vrai a priori : c'est une seconde pétition de principes.

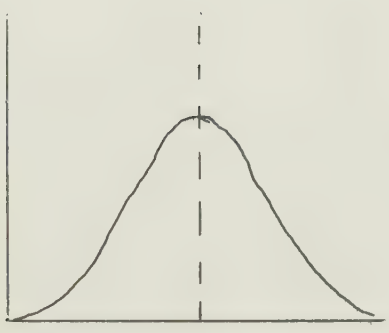
Il est fréquent que soient confondues médiane du possible et médiane du réel. Or, seule est définie a priori la médiane du possible (du correcteur). Celà revient à dire, non que le réel est possible mais que le possible est réel. C'est une troisième pétition de principes.

2) Le mythe de la courbe de GAUSS

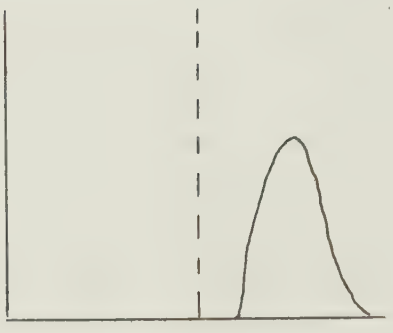
A priori, on estime généralement satisfaisante une distribution gaussienne de la population. Mathématiquement, la courbe ne dépend que de deux paramètres : la variance et la moyenne. Ce qui la caractérise c'est que la valeur moyenne (m) représente à la fois la moyenne, le mode et la médiane de la distribution (médiane du réel). Ainsi sont confondues moyenne de la population et médiane du réel.

Si la médiane du possible (celle du correcteur : 10/20) se confond comme c'est habituel, avec la médiane de la population, une distribution gaussienne entraîne, ipso facto, une confusion entre

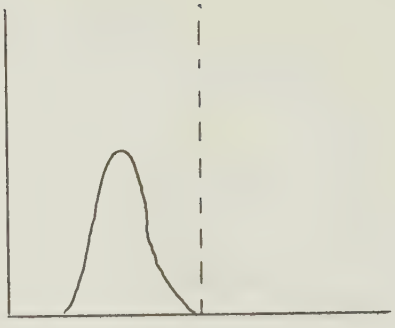
DEUX MYTHES



Une idée fixe



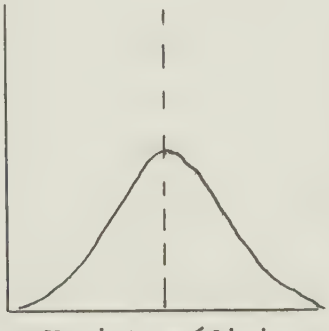
Un caprice (gentil)



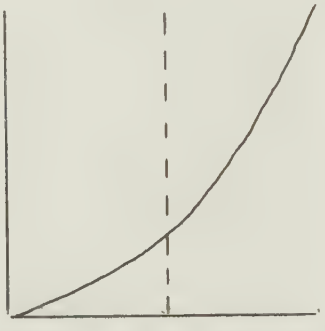
Un caprice (méchant)

Deux caprices de correcteurs

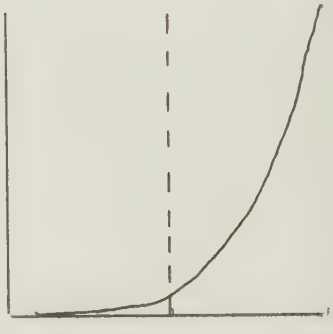
Un enseignement inexistant si l'examen est fiable.



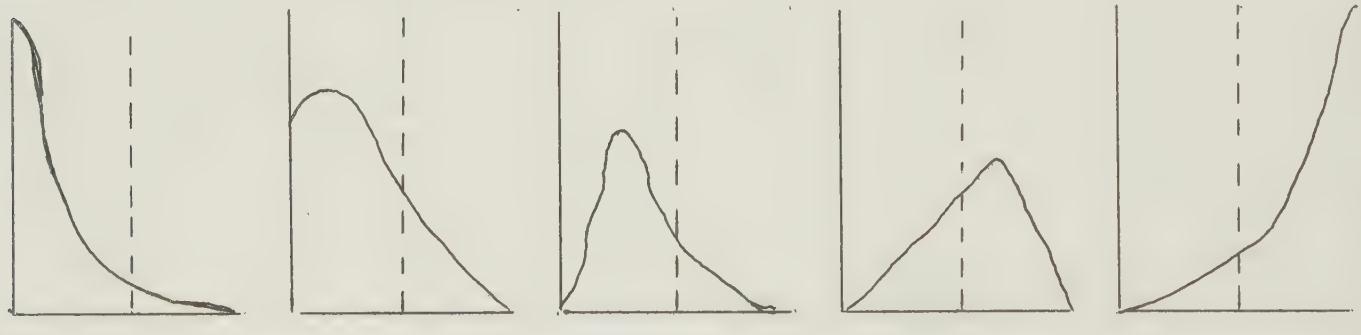
Un intermédiaire ou un caprice ?.



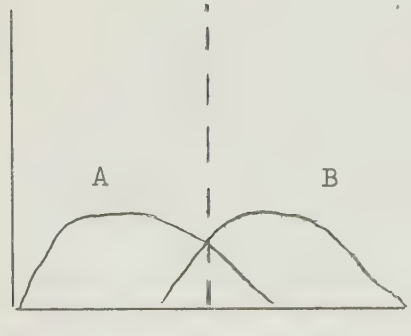
Passer au suivant serait un progrès.



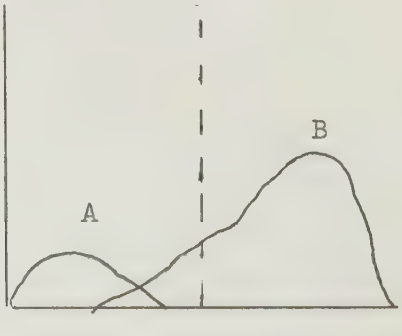
Un enseignement efficace si l'examen est fiable.



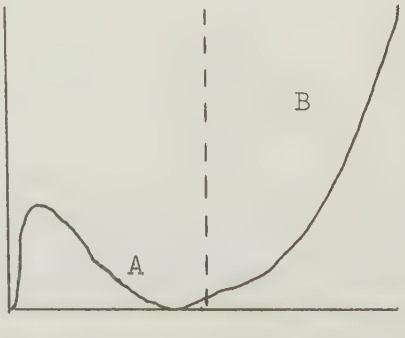
Une progression que ne montrera jamais le mythe gaussien



A progressera-t-il ?



OUI



NON

En pointillés la médiane

médiane du possible et médiane du réel puisque cette médiane de la population se confond elle même avec la moyenne. Comme il est par ailleurs constant que soient confondus seuil de suffisance (10/20) et médiane du possible (du correcteur : 10/20), LA CONFUSION EST TOTALE.

3) Deux mythes ensemble

Il est habituel de constater que la moyenne de la population ne s'écarte pas significativement de 10/20. Cette habitude implique qu'un tel fait soit accepté et donc considéré comme acceptable. Sinon les examinateurs ne sont pas logiques. Ajouter à ceci le caractère souhaitable de la distribution gaussienne revient à admettre :

a/ qu'il est souhaitable que la moyenne n'ait aucune signification, ce qui ne l'empêche pas d'être la note "suffisante"

b/ qu'il est souhaitable que l'effectif acceptable détermine la note et non que la valeur de la population détermine le nombre des reçus : c'est la PETITION DE PRINCIPES CAPITALE.

Comme il est par ailleurs constant que la médiane de la distribution possible des notes par le correcteur se confonde avec le "seuil de suffisance" (10/20) une distribution gaussienne autour du seuil de suffisance comporte quatre conséquences :

a/ 50 p. 100 de la population doit être refusée systématiquement. Celà signifie que, si l'on souhaite 100 étudiants en fin de 6ème année, il en faut 200 en fin de 5ème, 400 en fin de 4ème, 800 en fin de 3ème, 1 600 en fin de 2ème, 3 200 en fin de 1ère et 6 400 à l'entrée ! Ce qui représenterait un rendement estimé comme normal de 1/64 soit 1,5 %.

b/ arrêter ce processus "au bout d'un certain temps" relève du pur arbitraire car un examen où tout le monde est reçu systématiquement ne prouve rien et c'est bien ainsi que celà finit alors que c'est d'une autre manière (gaussienne ou proche) que celà commence.

c/ l'enseignement ne sert à rien puisque la population reste distribuée au hasard autour d'une "moyenne-médiane-seuil de suffisance" non pas laissée à la discrétion du correcteur mais dont la signification est laissée à la discrétion du correcteur.

d/ l'examen ne sert à rien puisque le résultat est une répartition unimodale alors que ses résultats sont binaires. Il s'avère incapable, dans un grand nombre de cas, de séparer les suffisants des insuffisants : il n'assume pas sa fonction de ségrégation.

VANITE DE LA STATISTIQUE

D. SCHWARTZ, dans son ouvrage cité en référence, écrit (p. 12) "(...) Malheureusement le mot prête à confusion : les statistiques, historiquement comme étymologiquement, sont depuis longtemps des dénombrements fournissant aux Etats des renseignements d'ordre administratif ; la statistique est tout autre chose : une méthode de raisonnement permettant d'interpréter le genre de données très particulières qu'on rencontre notamment dans les sciences de la vie, dont le caractère essentiel est la variabilité".

Les résultats d'examen qui ont fourni le point de départ à cette réflexion ont été abordés comme il est habituel de le faire. Des effectifs (des "etats") ont été examinés parce qu'ils expriment quelque chose de concret : un nombre d'étudiants ayant satisfait leur Professeur. Tout comme le personnage de comédie dont se gausse D. SCHWARTZ, c'est ainsi que le jury connaissait son monde mais avec plus d'un doute sur treize mille quatre cent quatre-vingt-dix-huit ! Un essai de comparaison a été tenté entre les cinq épreuves qui constituaient l'examen : à l'évidence, ces effectifs étaient différents. Tout au plus a-t-il été remarqué que, dans nombre de cas, un refusé dans une discipline avait une chance sur deux d'être reçu dans une autre : jouer le sort d'un tel refusé à pile ou face n'aurait pas changé grand chose et il ne semble pas que cette constatation soit encourageante. La disparité, dûment constatée et généralement significative entre les fractions de réussite d'une épreuve à l'autre posait deux questions. La première étant "telle discipline, à faible réussite, est-elle difficile, inabordable pour la plupart des étudiants, mal enseignée, trop sévèrement corrigée, mal explorée ... ?" Il n'est pas possible d'y répondre faute d'informations. Quant à la seconde "y a-t-il un rapport entre les notes attribuées et les effectifs reçus ?" Elle semble bien du domaine de la statistique.

La première surprise fournie par l'étude des notes, en les abordant par le chemin habituel en matière d'examen, ce fut de constater que leurs moyennes étaient proches, voire non significativement différentes. La question se posait alors des rapports qui existent entre des moyennes quasi identiques et des effectifs franchement différents si on rappelle que c'est, en principe, les notes qui fournissent le résultat. Il y avait, d'autre part, quelque chose de gênant à constater que la moyenne réelle se confondait plus ou moins avec la médiane. Mais la moyenne n'est pas tout. Sachant que très peu de candidats (moins de 20 p. 100) avaient été satisfaisants partout, que la plus grande fraction (26 p. 100 environ) avait été insuffisante dans deux épreuves sur cinq, l'hypothèse qui venait à l'esprit c'est que les étudiants "jouaient le système" leur permettant de "choisir" ce sur quoi ils allaient faire leurs points en fonction probablement de leur goûts ou de leurs aptitudes. S'il en était bien ainsi la corrélation entre les notes d'épreuves seraient faible. Or, elle est excellente ! On peut donc en déduire que la population est assez homogène et que la notation se fait de façon comparable d'une discipline à l'autre. Comment, dans ce cas, expliquer la disparité des effectifs satisfaisants : en 1970 telle discipline reçoit plus de 80 p. 100 de l'effectif et telle autre 31 p. 100.

A l'inverse de ce qu'on pourrait croire, la corrélation entre les notes des candidats reçus partout est faible, parfois nulle ou négative ! Il y a donc un certain désaccord entre le "caractère reçu" et la note qui permet justement de déterminer ce caractère.

A la limite du découragement, il restait à aborder ce qui représente le fond même de la statistique : la variabilité. Bien vite, il a été constaté que, non seulement les variances sont variables ce qui, après tout, n'avait rien de trop surprenant, mais que les variations de variances n'avaient pas la même signification partout. Une étude de la liaison des caractères aurait dû être faite mais il y a été renoncé : le considérable travail supplémentaire que celà aurait demandé a été estimé vain puisque, dans chaque discipline, c'est la grandeur quantitative qui définit le caractère qualitatif et que celui-ci définit, en fait sinon en droit, une certaine grandeur : 10/20.

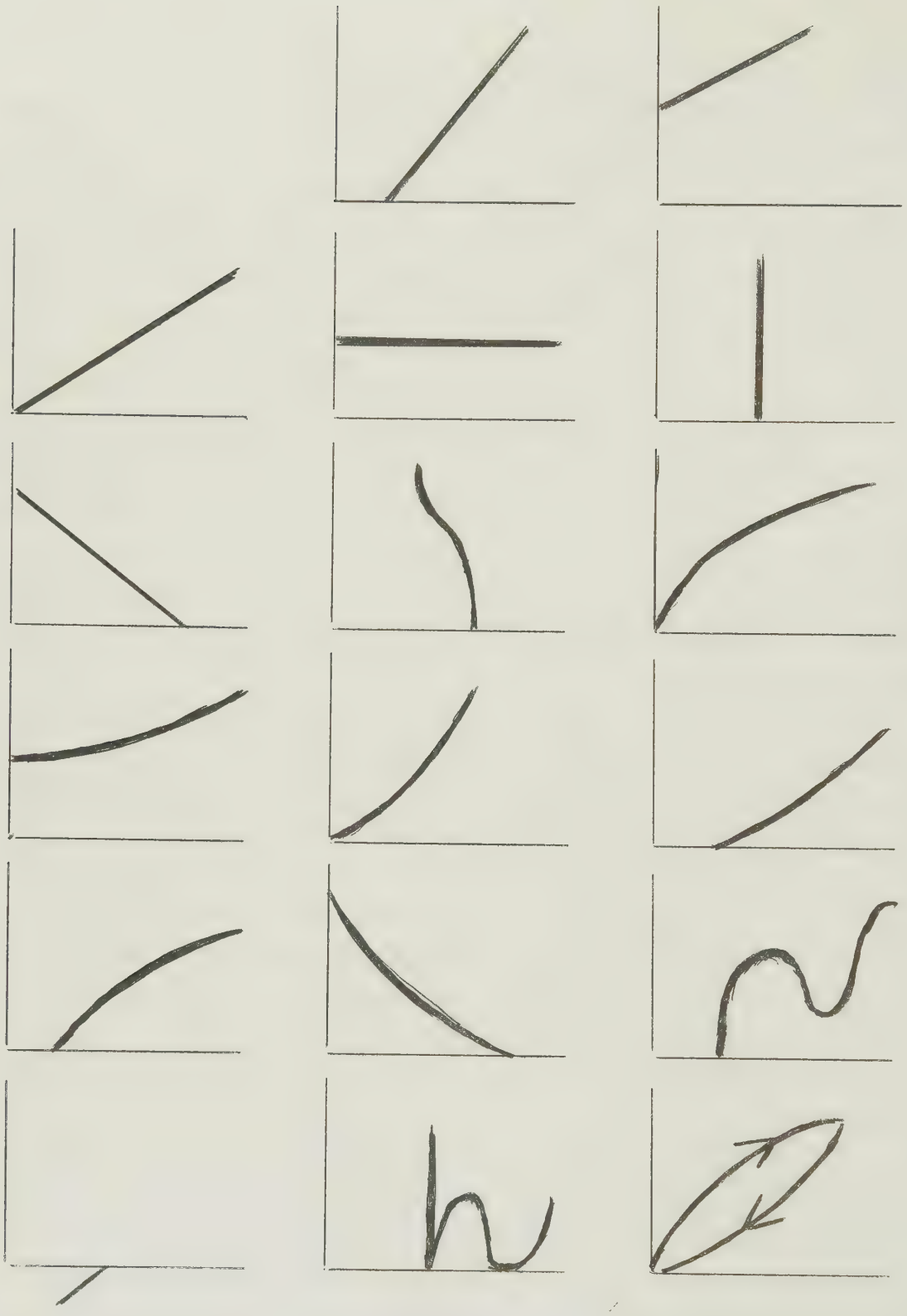
La méthode statistique semble inapplicable ou, en tout cas, inexploitable. La variabilité n'est appréciable que sur des étalons définis. Une durée peut s'exprimer en années, en semaines, heures ou

secondes et chacune de ces expressions peut aisément se transformer dans une autre. On peut établir des corrélations entre des tailles exprimées en pouces, en lignes, en yards, en mètres ou en coudées car chacune de ces unités est définie et les erreurs dans les résultats sont des erreurs de mesure. On peut compter le nombre de loges capsulaires dans chaque coquelicot, le nombre de descendants de chaque femme, le nombre de poils par centimètre carré de telle partie du corps : tout observateur ne trouvera peut être pas le même nombre mais pour tout observateur, une loge capsulaire, un descendant, un poil, un centimètre carré, une certaine partie du corps représentent des entités parfaitement définies même si la définition est arbitraire. Pour un examinateur, dans l'actuel état des choses, un point ne représente rien de défini. Le seul résultat des calculs faits, fut de montrer l'incohérence fondamentale du système puisque la base de tous calculs est fausse : l'unité des comptes qui représente une entité différente d'un observateur à l'autre et non définie pour chaque observateur.

Le caractère "reçu" n'est pas défini par un critère qualitatif mais par un nombre de points dont on ne sait pas ce que chacun d'eux représente. Qu'il suffise, par exemple, d'avoir cinq pièces pour être satisfaisant : on peut avoir une lire, un franc, une livre, un dollar et un Deutsch Mark aussi bien que cinq lires ou cinq livres. Comparer ainsi la variabilité de plusieurs sommes n'est possible que dans la mesure où on sait de quoi chacune est faite. On ne le sait pas et d'une étude ayant les formes de la statistique on peut tirer tout ce qu'on veut puisque les différences individuelles viennent en réalité non des individus observés mais de l'observateur et que celui-ci n'en a pas conscience.

Un jour il fut demandé à un jeune instituteur pourquoi cette copie "valait" 8.5 et non 8 ou 9. La réponse fut admirable : "mon père était déjà dans l'enseignement". Aussi longtemps que l'existence d'un gène docimologique n'aura pas été prouvé, aussi longtemps qu'on fera des calculs à partir de nombre ne représentant rien de défini, les statistiques en matière d'examen mériteront leur réputation traditionnelle de source infinie de mensonges ou d'erreurs. Il sera mieux de ne pas en faire afin d'éviter que ne se fixe définitivement un système déjà figé et que chacun y trouve argument pour persister dans ses erreurs.

UNE QUESTION, 17 + 1 REPONSES



QUESTION: Sachant que $\dot{V} = V_T \times f - V_D \times f$, représentez l'allure général des variations de $\dot{V}$ en fonction de V_T en supposant f et V_D constants.

LA NOTE N'EST PAS TOUT

Les documents étudiés ont permis de poser l'hypothèse que les "mesures" étaient faites avec un mètre en caoutchouc mais dans la plupart des cas, ils n'étaient pas d'une telle nature qu'il ait été possible d'en déduire qu'on mesurait n'importe quoi. Cependant, quelques observations privilégiées donnent à penser.

Dans une même épreuve, se tenant le même jour et concernant les mêmes candidats, deux questions, parmi quatre-vingts autres, ont été posées. L'une concernait l'intégrale d'une question complexe : elle a été réussie par soixante dix pour cent de la population. L'autre concernait une fonction du premier degré : soixante dix pour cent l'ont manquée. La question qu'on aurait pu prévoir difficile sortait du cours avec une telle force d'évidence qu'il n'y avait pas un couloir sans un groupe de candidats se rafraîchissant la mémoire à son sujet. On pouvait prévoir qu'elle était difficile, on a pu constater qu'elle était facile, elle était attendue et "sue par coeur". L'autre, facile d'aspect, s'est avérée difficile en réalité : elle n'était pas attendue sous cette forme. Alors que le cours avait enseigné "y=ax+b", la question était de la forme : en supposant b constant et x constant, que devient y lorsque a croît ? La réponse devait être fournie sous forme de graphique. La question réussie était une récitation, la question manquée faisait appel à la compréhension. Quant à se faire une idée du niveau de culture mathématique des candidats ce n'est possible que si on considére séparément chacune des deux questions. Malheureusement, les résultats sont en parfaite discordance. Une dernière précision : la fonction qui aurait dû être linéaire et ne passant pas par l'origine a été représentée de 17 manières différentes ce qui, avec les "questions blanches" forme un total de 18 versions (Représentation jointe).

CONCLUSIONS

On ne sait pas, en réalité, ce qu'exprime la note dans sa grandeur mais pas davantage, et dans nombre de cas, ce qu'explorent les questions ni si l'épreuve présente un rapport clair avec l'enseignement ni si l'enseignement est lui-même en rapport avec le but avoué des études. Même avec une notation parfaite, les résultats d'un examen resteraient douteux pour qui ne connaîtrait ni la forme ni le fond du procés. Or, les juges eux-mêmes ignorent ces faits puisque la plupart de ceux qui acceptent de "contrôler les connaissances" ne savent ni ce que signifie "contrôle" ni ce que sont les "connaissances". En plus, le verdict prend la forme de points absolument non définis. Tout se passe comme si des juges infligeaient un emprisonnement, exprimé en nombre de jours d'une planète imaginaire et différemment imaginée par chacun d'eux.

Le mal, c'est l'incapacité de discerner dans les résultats ce qui revient aux candidats, à l'enseignement ou à l'examen lui-même, d'engager gravement l'avenir à partir d'épreuves dont on ne sait ni ce qu'elles doivent prouver ni comment elles pourraient le prouver ; C'est d'accorder un pouvoir exorbitant à un système qui n'est pas fiable et qui ne peut l'être aussi longtemps qu'on ne saura pas ce qu'on en attend, sur quels critères émettre un jugement, comment exprimer ce jugement et, même, quand et quoi juger.

2ème PARTIE

UN REMEDE

Le mal c'est la confusion. Pour tenter de porter un remède à ce mal il semble qu'il soit avant tout nécessaire de s'entendre sur le sens des mots employés afin que se dégage une idée aussi claire que possible de ce que peut être chaque pas de la démarche d'un examen.

Ensuite, il sera envisagé une méthode permettant de rendre fiable le moyen de contrôle. En possession d'un instrument fidèle, sensible et sans dérive, il sera possible de le rendre juste, ou d'approcher cette justesse, en définissant la nature de ce qui doit être contrôlé.

Enfin, il deviendra possible de choisir le mode d'emploi de l'instrument. Après qu'aura été défini ce qu'on en attend on pourra choisir le moment et la forme de son utilisation.

Ainsi,cette seconde partie, qui se veut constructive, sera bâtie autour de cinq idées principales:

A- POSSEDER UN VOCABULAIRE PRECIS

B- DONNER UN SENS AUX RESULTATS ET UNE SIGNIFICATION AUX NOTES.

C- DEFINIR LES OBJECTIFS.

D- ETABLIR UN PROGRAMME.

E- CHOISIR UNE FORME ADEQUATE.

Eventuellement, chacune de ces idées se trouvera décomposée en chapitres.

CHAPITRE I

LES MOTS

A- POSSEDER UN VOCABULAIRE PRECIS

" Les mots sont sources de malentendus."

APTITUDE - CAPACITE 151
ARBITRAIRE - JURY - LOGIQUE 152
CANDIDAT - CORRECTEUR 154
CONTRÔLE 155
COURS - CURSUS - CURRICULUM 157
IMPORTANT - INTERESSANT 159
ITEM - PROPOSITIONS - QUESTIONS - EPREUVE 161
NOTATION - NOTE - NOTEUR 163
OBJECTIF 165
PEDAGOGIE - PEDAGOGIQUE 166
PROGRAMME 168
QUALITES : Parfait, nul, satisfaisant, stupide 171
SANCTIONS D'UN EXAMEN : Accepter, recevoir, refuser, numerus clausus..... 173
SCORE 174
CONCLUSION : EXAMEN 175

Pourquoi un triangle, en prenant du grade, devient-il quadrilatère puis polygone ? Le psychiâtre est-il à l'esprit ce que le jaunâtre est au jaune, une espèce d'altération ? (Variété serait-il plus juste ?). Un oculiste est réputé s'occuper des yeux (en latin) ; un ophtalmologiste sait en parler (en grec). Qui s'occupe de la vue ?

Le pédicure s'occupe des pieds : pourquoi pas l'orthopédiste ? (Comme les chaussures) Un orthopédiste d'adultes s'occupe-t-il de débiles ou de déments ? Qu'est-ce qu'un orthopédiste infantile ? Est-il immature ou bègue ? Que peut bien faire une intervention orthopédique sur la hanche d'un vieillard au psychisme intact ?

Donner le jour à un enfant c'est enfanter. "Cultiver" un enfant, c'est faire de la puériculture, mais donner la culture aux enfants n'est pas le fait d'un puériculteur. Redresser un enfant c'est faire de l'orthopédie, cependant, les défuntes "maisons de redressement" n'avaient pas l'orthopédie dans leurs attributions.

A toutes ces subtilités, dont notre langue est riche, déconcertante pour peu qu'on y réfléchisse s'ajoutent des mots à double sens, au moins, dont nous sommes friands : les mots d'origine étrangère, pour lesquels nous éprouvons une affection particulière surtout si, strictement homographes, homonymes, trop souvent, de leurs homologues français, ils ont une signification différente.

Il semble que, bien souvent, l'usage de termes imparfaitement définis soit pour quelque chose dans la confusion des résultats d'examen et, sans doute, de ses causes : l'enseignement, les questions .... et les notes.

APTITUDE - CAPACITE

Etes-vous capable de piloter un avion ? Oui, si vous êtes en état de répondre à la question après l'avoir fait. Si vous n'avez jamais piloté, vous devez répondre non. Ce qui ne préjuge en rien de vos aptitudes. Etre capable, c'est être apte. On peut être incapable et apte ou inapte.

Aptitude et capacité sont deux directions possibles pour un examen. L'aptitude est une disposition naturelle pour quelque chose. La capacité est une habileté ou une faculté légale. Littré parle de la "contenance d'une chose". Il semble que, dans le domaine de l'enseignement, "contenance" doive être remplacé par "contenu". La capacité légale, l'habileté et, a fortiori, le "contenu" (au sens figuré), sont des acquisitions alors que les aptitudes sont innées. Les acquisitions sont la finalité de l'enseignement, ce qui ne signifie pas que les aptitudes ne puissent être augmentées, précisément par l'augmentation du contenu et de l'habileté.

A la limite, il est vain de tenter de concevoir un examen purement d'aptitude ou purement de capacité. Cependant, un cas est bien clair : celui d'une capacité professionnelle, légalement définie, pour laquelle le candidat doit apporter la preuve qu'il sait faire le métier qu'il prétend exercer.

Au niveau scolaire, non terminal, il est réellement impossible d'examiner les aptitudes en dehors de toute acquisition. Cette impossibilité n'est peut être pas éternelle, mais elle est.

ARBITRAIRE (JURY - LOGIQUE)

Injure infamante souvent adressée à un jury d'examen. Qu'un jury prenne une décision arbitraire c'est un fait, ce n'est pas une injure : c'est sa fonction. Ce qu'on peut lui reprocher, c'est de ne pas le faire assez souvent.

Est arbitraire, ce qui est laissé à la discrétion des juges (arbiter = juge) ; ce qui n'a de règles que la volonté du prince, dit Littré. Un "correcteur" (un noteur) ne doit pas donner de décision arbitraire. Simple copiste, il doit fidèlement exécuter la loi (dura lex, sed lex) même, et surtout, si lui-même l'a édictée : il est alors logique. L'établissement d'un score, l'évaluation et la notation doivent suivre une méthode d'enchaînements rigoureux.

Le jury d'examen, dans la loi française, attribue la note. C'est un point difficile. Dans la réalité, il ne l'attribue pas, il l'authentifie si on prend le mot "note" dans son sens traditionnel d'expression chiffrée. C'est alors un organisme officiel dont la fonction consiste à apposer une signature sur un registre : c'est un contrôle. Dans l'acception la plus étroite du mot "note", le jury désigne (marque) ceux qui sont reçus. Son jugement est alors binaire et non plus chiffré. Dans ces conditions sa décision est arbitraire. En effet, déclarer reçu un candidat qui se présente devant le jury avec 15/20 dans chacune des épreuves constituant l'examen est une pure formalité que n'importe quelle machine ferait aussi bien que le jury, plus vite et avec moins de chance d'erreurs. La position est la même pour un quintuple zéro. En revanche, il est des cas litigieux qui nécessitent réflexion, au sujet desquels la décision n'est pas "automatique". C'est là que le jury est utile, indispensable. Attribuant légalement les notes, son pouvoir est discrétionnaire. Il a parfaitement le droit de juger qu'un candidat ayant obtenu 49,5 sur 100 doit être reçu : une machine ne l'aurait pas fait. Il est regrettable que, ce genre de décision prise, il perde un temps précieux à savoir dans quelle discipline il serait bien possible d'ajouter le demi point : ce n'est plus de l'arbitraire, c'est du formalisme.

Formé de gens réputés intelligents, un jury peut parfaitement estimer qu'une certaine note, par comparaison, à la suite d'une

analyse, ne semble pas indiquer une certitude d'échec. Il a alors le droit de recevoir un candidat après critique de la méthode. Un candidat ainsi reçu, l'est à la suite d'une décision arbitraire.... Et généralement il ne s'en plaint pas.

Si le jury ne prend pas plus de décisions arbitraires c'est, semble-t-il, bien souvent, parce qu'il n'ose pas. Pour sa défense, il est honnête d'ajouter qu'il est presque toujours mal informé, mais il est rare qu'il demande à l'être comme il en a et le pouvoir, et le devoir.

CANDIDAT ET CORRECTEUR

C'est un duo qui établit des rapports médiats.

CANDIDAT : C'est celui qui, vêtu de la toge blanchie, postule .... Voici un sujet qui se présente vêtu d'une manière particulière : on retrouve deux traits de l'examen traditionnel : la formalité et la circonstance rare. Cet aspect : démarche solennelle, forme rituelle, oriente plus vers la vérification et l'impression d'une estampille (impétrant) que vers un "contrôle" à l'anglo-saxonne. Dans cette perspective, un candidat n'est pas un étudiant (qui étudie) mais quelqu'un qui prétend ne plus l'être, au moins de la même manière.

LE CORRECTEUR est celui qui corrige. Corriger a trois sens apparents : ôter un défaut, modifier (correcteur de goût), infliger une punition corporelle. En réalité, ces trois actes ont une ambition commune : améliorer. Voilà qui contredit singulièrement la notion d'examen traditionnel. Les observations du correcteur (s'il les faisait) ne serviraient à rien, elles ne peuvent pas améliorer : elles arrivent trop tard.

Ou bien on accepte l'examen, vérification solennelle opérée sur des candidats, et le correcteur n'a rien à faire : c'est un "noteur" qu'il faut.

Ou bien on traduit "examiner" par "to control" et le correcteur est indispensable. Le contrôle en question est alors une opération d'enseignement.

CONTROLE

Il est devenu de bon ton de substituer au terme "examen" l'expression "contrôle des connaissances". Il semble que, en France, ce terme ait été introduit aux environs de 1960 : il se rapportait alors à une technique particulière de tests étalonnés. Ces tests ayant été employés antérieurement en Grande-Bretagne et aux Etats-Unis, il ne semble pas inutile de se poser la question d'une éventuelle différence de sens dans les deux langues.

En français, ce terme d'origine française, a un double sens : c'est une vérification et la marque attestant que le contrôle a été fait. Ainsi, on vérifie qu'un objet est en or (contrôle) et on atteste qu'il est bien en or en imprimant une marque avec un poinçon (contrôle). Le douanier contrôle les bagages des voyageurs et atteste que le contrôle a été fait en marquant le bagage contrôlé.

A l'origine, le contrôle était un registre (une liste) tenu en double. Cette notion de "double" se trouve dans maints autres mots (contremarque, contre-courbe, contredanse, contrepoint ... etc). L'idée de "liste" se retrouve dans rôle d'équipage, rôle d'impôts ... etc. "Rôle" vient de rouleau (de parchemin) rouler, roue (rota) et les mots anglais "roll", "to roll" ont bien le même sens et la même origine tandis que le doublet juridique "rôle" a même orthographe, presque même prononciation et rigoureusement le même sens.

Mais si on peut traduire contrôler par vérifier, "to control" se traduit par "diriger", "régler" et la différence est considérable. Si on admet que, pour un français, le thermomètre est un contrôleur (il vérifie la température) pour un anglais, un contrôleur de température ne sera pas un thermomètre mais un thermostat. Par ce biais, il est possible d'attribuer à l'examen contrôle (contrôle des connaissances), deux fonctions :

- une fonction "française" de vérification et de marque (note) de cette vérification : il est juge inactif.
- une fonction "anglaise" de direction : il est juge actif.

Ce n'est pas tant l'examen qui est corrigé : c'est lui qui corrige ou qui peut corriger, c'est-à-dire améliorer.

Il est évident que ces deux attributions peuvent parfaitement se compléter. Si, tout en conservant à l'examen son caractère d'estampille, on lui ajoute une fonction corrective, il semble bien qu'il devienne autre chose qu'un moyen de se rendre compte (cf examen) :

un acte parfaitement intégré à l'enseignement. Un acte d'enseignement. Encore faudrait-il qu'il ne vînt pas systématiquement lorsque tout enseignement est terminé. Le terme de "contrôle continu" prend alors un sens tel qu'il est presque un pléonasme.

COURS - CURSUS - CURRICULUM

Dans une perspective d'activité d'enseignement, trois définitions du mot cours méritent méditation.

1°- Le cours de l'eau. N'est-ce pas suivre le cours de l'eau, se faire porter par la routine ? Littré propose un sens un peu plus actif : un cheminement dans le sens de progrès. Contrairement à son aspect routinier, le cours d'eau présente une notion de "force des choses" : on ne fait pas remonter son cours à un fleuve. Dans l'enseignement, il y a un amont et un aval. Un certain dynamisme intellectuel, des étudiants ou des écoliers, se développe toujours dans le même sens et ce serait un vain gaspillage d'énergie que de tenter de renverser cette force. Elle existe, il faut bien en tenir compte ; elle a un sens : il faut le connaître. On peut l'endiguer, la dévier, non la supprimer. Par exemple, c'est aller dans le sens naturel des choses que de procéder du concret vers l'abstrait et du simple vers le complexe. A un niveau d'études relativement élevé : celui du premier cycle des études médicales, un enseignement est étonnant : celui de l'Anatomie. Certes, il est ce qu'il est depuis longtemps (routine ?) parmi ceux qui l'ont subi, il en est qui en savent encore quelque chose (mais ne l'ont-ils pas réellement appris plus tard ?) Est-il réellement logique de commencer par apprendre ce qui ne se voit pas pour finir par ce qui se voit ? Encore ce dernier objectif est-il rarement atteint ! Ne serait-il pas plus logique de se servir du fil de l'eau : de la curiosité et du sens du concret pour progresser vers une connaissance qui est une abstraction pour la majorité des médecins (qui ne sont pas tous chirurgiens) ? La géométrie des classes élémentaires ne doit-elle pas être d'abord une opération concrète quant aux formes, aux proportions, puis, ensuite seulement, un formulaire et une séquence de théorèmes ? Après tout, c'est ainsi qu'elle a été inventée et ce n'est sans doute pas pour rien que les géomètres de l'antiquité ont été aussi physiciens et astronomes : cette dernière discipline restant, même au XX° siècle dans sa seconde moitié, largement abstraite.

2°- Un ensemble de leçons sur un même sujet

Le cours magistral est mal considéré : il est devenu exposé, ce qui suppose que le maître est resté et la cohérence partie. C'est

dénaturer le sens du mot, et de l'acte, que de dire que Monsieur Untel fait son cours tel jour à telle heure. En général, un cours dure plusieurs heures et s'étend sur plusieurs jours ou semaines. L'ensemble, supposé par le mot, implique une progression dans l'exposé : il part d'un point et aboutit à un but. Encore faut-il que, pour l'enseignant, au moins, le but soit défini, sans quoi son cours prend un troisième sens.

3°- Une promenade plantée d'arbres

Où poussent sans doute les fleurs de rhétorique. On se promène un certain temps (on emploie son temps) mais une promenade n'est pas un parcours.

CURSUS est parfois employé au lieu de cours. C'est un mot latin, c'est évidemment plus élégant, à moins que ce ne soit pédant. Mais sa signification semble assez éloignée de ce qu'on prétend lui faire dire : c'est la succession des magistratures, la suite (carrière) des honneurs dans son sens originel. Puisque le mot est latin, et non français, son sens est latin, éloigné, très éloigné d'un ensemble de leçons faites sur un même sujet. Littérairement, il désigne des cadences harmonieuses terminant une période. Carrière des honneurs ou cadences harmonieuses sont plus le souci des professeurs que l'objectif des étudiants pris en tant que tels. Cursus est un terme qui ne signifie pas cours. Il est, dans cet emploi, pédant.

CURRICULUM est, en revanche, parfaitement acceptable. Ce terme employé notamment par les anglo-saxons signifie "carrière" qui vient du bas latin, presque de l'italien "via carraria" : chemin de chars. Nous dirions aujourd'hui "autoroute". Or, une route est faite pour aller d'un point à un autre et elle peut être adaptée à certains usagers. Le curriculum c'est donc ce qui mène les étudiants d'un certain degré d'ignorance (parfois relative) à un certain degré de savoir. C'est un itinéraire, un cours organisé, presque un programme.

Au total,

Un bon cours est ce qu'on peut souhaiter aux étudiants.
Un bon cursus, ce qu'on peut souhaiter aux enseignants.
Un bon curriculum, ce qui rendrait service aux deux si la route était adaptée, orientée : tracée et non "en projet".

IMPORTANT - INTERESSANT

Si on admet qu'est "important" tout ce qui a ou peut avoir de grandes conséquences, un examen est important. Chacune des épreuves qui le constituent et chacune des "questions" qui les composent sont donc importantes. Un examen est important, les questions sont relativement importantes et cette relativité dépend de leur nombre. Une épreuve dépend en fait, d'un Professeur ou d'une équipe d'enseignants puisqu'il est habituel qu'elle soit composée par ceux-là mêmes qui ont fait l'enseignement. A ce titre, toute épreuve a l'ambition d'être importante, d'avoir de l'influence sur le résultat final. Certains ne réussissent qu'à affecter d'avoir de l'influence car ici comme ailleurs, il est possible de se contenter des apparences et de confondre ambition et vanité. Ainsi, tout examen est important : pour les candidats(importance objective) et pour les examinateurs(importance subjective). Le point essentiel est de savoir si la seule vanité des uns est ou non déterminante des conséquences pour les autres.

On confond souvent "important" et "intéressant" et on a raison puisque "interest" signifie "qui importe". Mais il faut encore distinguer ce qui intéresse de ce qui est digne d'intérêt et, les deux ne sont pas nécessairement liés. En termes d'enseignement, n'importe quoi peut intéresser dans la mesure où l'intérêt a été proclamé et, sur ce point, deux cas peuvent se présenter ; on a déclaré digne d'intérêt ce qui l'était ou ce qui ne l'était pas ou, inversement, pour l'indignité. Ainsi, le doyen de la faculté de médecine de Paris déclara que la circulation du sang était : "paradoxale, inutile à la médecine, fausse, impossible, inintelligible, absurde, nuisible à la vie de l'homme" (GUI PATIN, 1601-1672). On peut donc penser que la circulation, proclamée indigne d'intérêt s'est trouvée sans importance à un certain moment et dans un certain lieu : une découverte importante pour les grandes conséquences qu'elle pouvait avoir(et qu'elle a eues) était inintéressante. Déclarer un objet digne d'intérêt est l'une des fonctions de l'enseignant mais encore faut-il qu'une telle déclaration soit l'aboutissement d'une réflexion et non la manifestation d'un caprice.

Est intéressant ce qui comporte de l'intérêt. L'intérêt, en termes d'examen, c'est ce qui rapporte : une question cotée dix points est plus intéressante qu'une question cotée un point. En dehors de ce qui est avantageux, l'intérêt, c'est ce qui est utile et c'est encore ce qui fait rechercher le nécessaire, l'utile ou l'agréable. Lorsqu'une "question" est jugée présenter de l'intérêt, il faut encore savoir en quoi et c'est finalement sur un triple critère qu'elle doit être jugée.

"Nécessaire" ne se rapporte évidemment pas à la question elle-même, car la forme qu'elle suppose n'est pas obligatoirement la seule possible, mais au savoir qu'elle permet de révéler sur un sujet donné. Si le savoir révélé est nécessaire, la question est réellement importante et on peut encore prétendre qu'elle l'est si le sujet est "utile" à condition de savoir pour qui : pour le candidat, dans son présent d'étudiant ou son futur de professionnel ? Elle est importante. Pour le spécialiste qui pose la question et pour lui seul ? Elle ne l'est pas. S'il est possible que le candidat actuel devienne un tel spécialiste, il n'est pas sûr qu'il le devienne, il n'est pas mûr pour le devenir et, serait-il l'un ou l'autre, il resterait l'exception dans la population examinée.

Enfin, trop souvent semble-t-il, un sujet est déclaré présenter de l'intérêt parce qu'il est "agréable" pour l'examinateur. Si l'examinateur est aussi l'enseignant, l'intérêt en a été proclamé, ne serait-ce que par la grandeur du temps qui y a été consacré. Le professeur a oublié le temps parfois considérable qui lui a été à lui-même nécessaire pour découvrir l'intérêt du sujet et acquérir le savoir requis : exiger que ses étudiants accomplissent la même démarche, en un temps considérablement plus réduit, est peu réaliste. C'est, de plus, despotique dans la mesure où le temps réellement nécessaire est beaucoup plus long que le temps imparti et si l'intérêt est du seul domaine de la spécialité à laquelle il consacre son temps.

Au total, si on ne peut refuser à toute question le caractère 'important" qu'on lui attribue généralement, si on peut souhaiter que tous les sujets proposés dans une épreuve soient importants, il faut, avant tout, se demander "pour qui" ?

ITEM - PROPOSITIONS - QUESTIONS - EPREUVE

ITEM - Terme qui sera souvent utilisé. Il est toujours à prendre dans son sens strict d'unité comptable. Il arrive que soient superposées deux notions : item et examen de type mécanographique. Cette confusion est ici absolument hors de propos. Si l'examen est de type QCM il est possible que chaque proposition soit un item ou que plusieurs propositions forment, ensemble, un seul item. Dans une forme rédigée, un item, c'est ce qu'on cherche en lisant : la liste des item attendus constitue la "grille". La lecture consiste à cocher les item à mesure qu'ils sont découverts et acceptés. Ainsi, l'item est l'unité comptable du score. Les rapports entre item, valeur et note seront examinés ailleurs.

QUESTION - C'est encore un terme à multiple sens. Son origine (quaerere) indique sa fonction d'enquête, de recherche.

L'usage a consacré un sens interrogatif. Une épreuve d'examen s'appelle souvent un questionnaire et une question est de la forme "que savez-vous de .... " "quels sont les .... " "que pensez-vous de .... " etc. Une phrase interrogative considérée comme sujet d'épreuve peut revêtir deux formes :

- ouverte : elle se termine par un point d'interrogation,
- fermée : après le point d'interrogation sont suggérées des réponses entre lesquelles il s'agit de faire un choix.

Question a, en réalité, un sens plus large : c'est une demande à l'effet de s'informer de quelque chose, c'est une proposition à examiner, à discuter. Le célèbre "that is the question" n'a pas le sens interrogatif. Pour éviter toute confusion, le terme souvent utilisé sera :

PROPOSITION - Littré lui attribue exactement le même sens : action de proposer (proponere : mettre devant, présenter), de soumettre à un examen, à une délibération. La forme n'est pas précisée. Une question est une proposition interrogative ; une proposition peut encore être affirmative ou négative. Le fond d'une affirmation (ou d'une négation) peut être, à volonté, vrai ou faux alors qu'il est difficile de poser une question fausse, échappant à l'ambiguité, et à laquelle il ne soit pas très difficile de répondre.

EPREUVE - C'est la matière de l'examen. Considérée d'un point de vue comptable, c'est un ensemble d'item. Elle est destinée à permettre au candidat de fournir la preuve qui montrera la réalité d'un fait : son savoir ou son ignorance. Une fois encore, il faut souligner qu'une épreuve mal constituée a peu de chance de manifester la preuve recherchée. C'est l'épreuve qui doit être le souci constant de l'examinateur. Constituer une épreuve est une opération beaucoup plus complexe que la noter, beaucoup plus redoutable et beaucoup moins redoutée.

NOTATION - NOTE - NOTEUR

Noter est souvent employé pour "corriger" et pour "évaluer". Or, ces trois mots désignent des actes fondamentalement différents. Noter, c'est mettre une note. Or, "nota", c'est un signe, une marque. Tout au plus, peut-on lui attribuer le sens : "commentaire à partir d'observations". C'est bien ce que signifie une observation d'un professeur sur le travail d'un élève. Le temps est passé où on écrit des discours sur les copies de ses élèves (mais pas pour les élèves). L'habitude est prise de représenter ses observations par des signes : c'est la notation. Que cette habitude soit regrettable ou non est une chose, mais c'en est une autre que de faire entrer dans des opérations des signes qui se trouvent être des chiffres (mais non des grandeurs) et cette habitude-ci n'est pas regrettable. Elle est absurde. Force est de conserver quelques habitudes de vocabulaire et admettons que la note soit l'ultime signe à fournir par l'examinateur à l'administration puisqu'elle l'exige. Mais alors, qu'après le mot "note", soit toujours, au moins sous-entendu, le mot "comptable", c'est-à-dire fournie à un service de comptabilité. Il s'agira bien d'un signe conventionnel fourni à un service centralisateur qui risquera sensiblement moins les absurdités si tous ceux qui s'adressent à lui parlent la même langue. Mais ce n'est pas parce que 10/20 aura pour tout le monde la même signification qu'on aura pour autant le droit de dire, ici, que deux fois dix font vingt.

Chacun sait que l'habitude est prise d'exprimer l'acuité visuelle en dixièmes. Il ne s'agit absolument pas de "dixièmes" de vision normale, mais de "pouvoir séparateur" exprimés en dixièmes de secondes d'arc. Une acuité de 15/10° n'étonne en général que celui à qui on l'annonce. Cette grandeur est réellement mesurable et définie par un étalon. Une acuité de 8/10° est encore fort honorable. On peut essayer d'expliquer à un borgne que, son oeil qui voit ayant une acuité de 16/10° il est, en moyenne, doué d'une assez bonne vision. Il n'en reste pas moins borgne et, tout malvoyant qu'il est, il comprendra parfaitement que certaines moyennes sont absurdes.

Si on accepte de "note" le sens que tout le monde y voit : l'expression d'une valeur, encore faut-il que cette valeur soit définie. Définie à l'avance et non reflet du sentiment qu'inspire la lecture au lecteur. Un ancien terme, archaïque : NOTEUR signifiait

"copieur de notes". C'est bien ainsi qu'on doit noter : en copiant, en d'autres termes, ce que montre la copie.

Au total, l'expression finale : la note comptable est l'aboutissement d'un processus complexe et le simplifier serait le déformer. Un stade intermédiaire s'impose : l'établissement d'un score, transformé en expression de valeur. Celui qui établit le score est un noteur.

OBJECTIF

C'est un éloge rarement décerné à un examinateur. Et pourtant, selon les points de vue, on peut dire d'un examinateur qu'il l'est souvent ou rarement ; mais il est à craindre que l'objectivité qu'on lui demande ne soit pas la meilleure possible. Tout dépend de l'objet.

La plupart des correcteurs tiennent compte des caractères de l'objet-candidat : c'est pourtant ce qu'on leur reproche de ne pas faire. Le défaut commun c'est de mal apprécier les dits caractères. Maints examinateurs jugent en fonction du candidat qu'ils imaginent et non de celui qui est. Cette erreur de jugement est souvent taxée de subjectivité et on a raison de penser ainsi. La malheur, c'est qu'il est extrêmement difficile de convaincre un examinateur, soi-disant objectif, et de bonne foi, que, ce qu'il croit être objectif, est subjectif. Peu d'examinateurs considèrent l'épreuve comme l'objet. Beaucoup d'épreuves étant mal construites, leurs résultats sont étonnants et l'examinateur désemparé se fixe sur ce qui lui reste alors : le candidat moyen .... et l'erreur de jugement qui s'ensuit.

C'est encore, dans le langage scholastique et cartésien, une entité intermédiaire entre le monde extérieur et la pensée. L'objectif, ainsi défini, est un concept, une création de l'esprit : les mauvaises langues parlent de vue de l'esprit. La vraie question est de savoir si l'entité est intermédiaire entre la pensée du concepteur et un monde réel ou imaginaire.

Enfin, un objectif, c'est un terme militaire que tout le monde comprend : un but à atteindre. Pour peu qu'il soit identifié, défini, et, même, publié, les chances de l'atteindre sont plus grandes que s'il reste vague, mystérieux, inconnu.

PEDAGOGIE - PEDAGOGIQUE

Les mots vivent et changent de sens. A l'origine le pédagogue était l'esclave qui menait les enfants à l'école. Comme à l'époque, il n'y avait ni café, ni tiercé et qu'il fallait bien employer son temps, pourquoi ne pas assister à la leçon du maître ? On peut supposer que ces accompagnateurs aimaient à briller en répondant à des questions difficiles que les marottes du maître (ont-ils changés ?) amenaient périodiquement. Au fil des ans, ils sont devenus "guides" en devenant répétiteurs. Mentor leur a servi de modèle : ce pédagogue qui avait brillamment réussi auprès d'Ulysse (et peut être, aussi, de Télémaque) est devenu ce que d'autres sont devenus sur ses pas : guide d'enfants non plus vers l'école mais dans l'école. Longtemps remplacés par le magister, ils ont été ressuscités lorsque l'art de mener les enfants a voulu devenir une science. Un interrègne a été assuré par les pédants qui n'ont pas laissé un très bon souvenir (on sait qu'un "pédant" était un "pedante" -italien- mais la filiation avec les pédagogues reste incertaine).

Quoiqu'il en soit, un pédagogue s'occupe d'enfants et, lui, contrairement à l'orthopédiste, ne l'a pas oublié. Montaigne avait peut-être été sage en écrivant l'institution des "jeunes gens". Que des instituteurs soient pédagogues, rien de choquant, au contraire mais, que des enseignants d'enseignement supérieur le soient, laisse rêveur.

Peut-être ne le sont-ils pas ? Le fait est que, s'ils le sont, ils ne le doivent pas à une acquisition scolaire puisque, légalement, aucune aptitude particulière n'est requise pour enseigner : c'est un exercice d'amateur (et encore n'aiment-ils pas tous cela).

S'ils sont sensés l'être, le mot désigne plus une qualité qu'une science. Mais alors, est-ce réellement une qualité que de considérer des adultes comme des enfants ? Il est généralement admis qu'une certaine jeunesse (l'adolescence) est significativement prolongée chez les étudiants, au sens universitaire du mot. Ce fait, mériterait qu'on s'y intéresse. Ou bien on prolonge chez des adultes un certain état d'enfance : d'immaturité et d'inaptitude à l'exercice de la liberté et cela n'est probablement pas un avantage. Ou bien, on prolonge, au-delà de l'âge habituel, une certaine plasticité d'esprit propre

à des acquisitions nouvelles et ce serait un avantage. Mais, à supposer que des méthodes soient appliquées, il ne semble pas qu'elles doivent être les mêmes de la maternelle au troisième cycle de l'enseignement supérieur. Et que dire de la pédagogie des adultes dans la formation permanente ?

Au total, et dans le cadre de l'enseignement traditionnel, scolaire et universitaire, il semblerait utile que soit faite une distinction de méthode en fonction de l'objet : enfant ou adulte jeune. A ce titre, il semblerait utile que "l'hébagogie" (1) se différencie de la pédagogie et que ceux qui ont à la connaître la connaisse.

---

(1) Néologisme ... K.B.

PROGRAMME

A l'origine, ce terme désigne un écrit qu'on affiche (publie) pour exposer le détail d'une fête publique, de la distribution des rôles, les conditions d'un concours. Dans toutes ces acceptions, se retrouvent les notions de catalogues et de "fête" ou de manifestation publique. Par extension, le programme est, justement, un manifeste, un exposé d'intentions : annonce du sommaire d'un cours, programme électoral, ...

Un autre sens, parfaitement défini mais d'apparition plus récente, est celui de l'informatique : le programme est l'ensemble des instructions fournies à une machine pour la résolution d'un problème. Mais que veut dire programme dans son sens scolaire ?

Il a certainement un sens de manifeste, de publication. Il est destiné à rendre publique quelque chose. Quelle chose ?

1° - On peut lui attribuer son sens théâtral. A l'origine, lorsque le théâtre existait, mais non l'imprimerie, on remettait à chaque comédien le texte (écrit sur un rouleau) qu'il devait apprendre. Ainsi, la distribution des rôles était une distribution de rouleaux (de textes) à apprendre (rôle se retrouve dans contrôle). Cette acception peut aussi bien se traduire par la répartition, rendue publique, de la tâche qui revient à chaque enseignant que par la distribution de ce que chacun doit apprendre. D'ailleurs un bon enseignant ne doit-il pas être un bon comédien ainsi qu'un bon candidat ?

2° - Son sens administratif : pour postuler tel grade il faut savoir telle chose.

3° - Son sens de manifeste : "moi, Untel, chargé de tel enseignement, je le présenterai de telle manière et à tels moments".

Le programme ainsi considéré avait bien un sens de catalogue : l'énumération d'articles rendue publique. L'exploitation de ce sens n'est pas inutile. Cependant, on peut reprocher à la plupart des programmes scolaires d'être rédigés en termes tellement vagues que chacun, ou presque, peut l'interpréter à sa manière. Dans la mesure où le programme, dans son sens habituel de catalogue de matières, est fait pour assurer une certaine unité d'exécution et qu'il permet diverses interprétations,

il n'est pas adapté à son but : il ne remplit pas sa fonction. Il risque de n'être plus qu'une forme vide de sens : il est alors plus une gêne qu'un aide.

Le sens moderne, qui n'exclut pas le sens classique, introduit une notion de logique et de chronologie. Cependant, il ne se superpose pas au terme habituel d'emploi du temps. On peut même avancer que ces deux expressions sont antinomiques. "Emploi du temps" signifie employer le temps. Cela suppose que soit d'abord déterminé le temps : par exemple, 39 semaines de tant de jours (donc de tant d'heures) à occuper. Au départ, les cases horaires sont vides et on les remplit. C'est le temps disponible qui impose une limite à l'enseignement.

"Programme" signifie ensemble d'opérations à faire. Chacune étant définie, à chacune peut être attribuée un temps raisonnable pour son exécution. La somme des temps étant calculée, se trouve, par les matières déterminé le temps nécessaire à leur enseignement. Ce n'est pas tout. Il reste l'aspect logique. Dans la mesure où telle discipline est jugée devoir être enseignée (donc "sue") il faut supposer que ce jugement est motivé. Par exemple, que l'une doit être enseignée avant l'autre parce que, dans la seconde, des notions nécessaires à son apprentissage doivent avoir été préalablement enseignées. On ne voit pas qu'il soit possible d'apprendre à calculer (non à compter) à des gens qui ne savent pas écrire les chiffres. L'étude de la littérature vient après l'apprentissage de la lecture .... etc. En allant plus loin, on peut attribuer plusieurs sens au terme scolaire de programme :

1° - C'est un catalogue d'objectifs. Il est souhaitable que ce catalogue, tout comme un autre, puisse être commodément feuilleté. Une organisation fondée sur les signes de lecture est logique : par exemple, l'ordre alphabétique. C'est un inventaire.

2° - Il est souhaitable qu'un tel catalogue, représentant une liste d'opérations rationnelles des enseignements soit rédigé en termes logiques et présenté en un ordre logique : taxinomique. Que la logique qui préside à la présentation d'objectifs d'enseignement soit une logique d'enseignement.

3° - Il est indispensable qu'à l'échelon, éventuellement, de l'unité d'enseignement la plus petite (classe par exemple) le temps imparti à chaque rubrique soit déterminé. Mais il est réaliste de considérer que,

tous les individus n'étant pas semblables entre eux, une marge d'erreurs dans l'appréciation du temps nécessaire soit prise en considération.

4° - Il est indispensable que les objectifs soient atteints dans un certain ordre dans la mesure où l'atteinte de l'un permet l'atteinte de l'autre. Ce qui est vrai pour une série, l'est encore pour un ensemble de séries, mais, incontestablement, une telle programmation n'est pas chose aisée et n'a, d'ailleurs, été que rarement tentée tant en France qu'à l'étranger.

Au total, deux sens peuvent être retenus : programme d'objectifs, programme d'enseignement. Ce sont deux choses différentes. Il semble utile que les objectifs d'enseignement soient, dans une large mesure, définis en termes identiques pour un très vaste ensemble de populations. Les besoins étant les mêmes pour un même objectif d'études dans tous les pays, les hommes étant semblables entre eux, intellectuellement parlant, au moins dans les pays de même civilisation, la définition de ces objectifs par les pouvoirs publics ou, de préférence, leur promulgation, n'a rien qui puisse choquer. A condition, bien entendu, que la rédaction soit claire, ce qui suppose que les fonctionnaires qui en sont chargés daignent s'informer avant, sachent de quoi ils parlent et n'aient d'autre souci que d'agir en fonction du but à atteindre.

Quant au programme d'enseignement, s'il est admissible qu'il soit défini en termes identiques pour tous, que le découpage temporel soit normatif, il ne peut pas être précisément prévu. En d'autres termes, les pouvoirs publics dans un but d'harmonisation peuvent prévoir la répartition dans de larges tranches de temps (qui s'expriment en mois), mais c'est au niveau de la plus petite unité d'enseignement, près de ceux qui, après tout, sont l'objet de l'activité, que peut être programmé le temps en heures.

QUALITES

Un mot les désigne toutes : perfection. "Je ne mets jamais ni zéro, ni vingt, disent certains examinateurs, car, chacun à leur manière, ils expriment la perfection et la perfection n'est pas de ce monde". Appréciable trait d'esprit, brillant paradoxe, commode excuse, éventuellement, à une notation "serrée" mais, dans la réalité, que peut-on en penser ?

PARFAIT : qui réunit toutes les qualités dit le sens commun. Littré est plus nuancé : est parfait, dit-il, une chose qui réunit toutes les qualités ; parfaite, une personne qui réunit beaucoup de qualités. Il ne dit pas combien. "Perficere" signifie "achever" et ni "parfaire", ni "perfectionner" ne signifient autre chose. Peut-on considérer "parfait" un candidat ? Oui, s'il réunit toutes les qualités qu'on attend de lui. C'est le niveau de l'épreuve qui déterminera celui de l'exigence et non la manière d'apprécier. Il n'est pas honnête de dire "je ne mets pas 20/20 à ce candidat. Certes, il a répondu de manière satisfaisante à toutes les questions posées, aucun reproche ne peut lui être adressé. Mais ce n'est pas avec le niveau de l'épreuve que je peux le déclarer génial". Il n'est pas demandé à une épreuve d'examen (contrôle des connaissances) de déceler le génie et, si le candidat avait répondu au-delà des questions, il eût été répréhensible de sortir du sujet. On peut, à la rigueur, et ce n'est pas certain, juger que 20/20 n'était pas "cher" mais il fallait concevoir l'épreuve autrement.

NUL, sans valeur. Dans un examen, il semble difficile de mettre zéro, sauf à des copies blanches. Une copie blanche pourtant, ne permet pas de dire que son auteur est nul mais de dire qu'il n'a pas apporté la preuve de la moindre connaissance. Une copie bien rédigée mais sur un autre sujet, n'est pas un indice de nullité de valeur : bien rédiger, ce n'est pas rien ! Le zéro est cependant justifié car la démonstration n'est pas celle qu'on attend. Ce qu'on attend d'un candidat, c'est qu'il manifeste non la preuve de connaissances, mais celle de certaines connaissances, en quantité suffisante sans erreur ou omission redhibitoire. Une quantité notoirement insuffisante (mais pas forcément nulle) peut entraîner un jugement de nullité "relatif" donc consolante.

SATISFAISANT : qui en fait assez. C'est encore une fois, le niveau de l'épreuve qui doit déterminer cette quantité.

STUPIDE : (engourdi) peut être qualifié de stupide un candidat qui manque de jugement. Pour cela il faut avoir su exiger de lui qu'il comprenne ce qui n'est pas nécessairement synonyme de connaître.

Ainsi un candidat juste satisfaisant pourra être dit moyen, un candidat qui montre tout savoir de ce qu'on lui demande : parfait ; un autre qui ne sait rien de ce qu'on lui demande réellement, nul. Il est tout à fait possible de déclarer parfait un candidat stupide si on ne lui demande pas de montrer qu'il a du jugement.

SANCTIONS D'UN EXAMEN

ACCEPTER, PRENDRE, RECEVOIR sont trois termes parfaitement synonymes. Parmi les 81 acceptions du mot "prendre", Littré propose : accepter, recevoir (n° 21). Cependant, pour éviter des périphrases, dans le texte qui précède, "accepté" a été choisi comme sanction des épreuves et "reçu" comme sanction de l'ensemble des épreuves constituant un examen. En termes réglementaires, "recevoir" a été pris comme l'une des deux sanctions du jury.

REFUSER est l'autre élément du couple de sanctions qui s'attachent à un examen. L'analyse de résultats réels semble montrer que refuser c'est, en fait, ne pas accepter. Les refusés sont les "laissés pour compte". C'est, semble-t-il, une grave erreur : refuser ne se définit pas par la négative mais d'une manière forte, précise et parfaitement en rapport avec les buts d'un examen : c'est récuser, déclarer incompétent ("refusare" venant probablement de "recusare") les deux termes français sont des doublets de signification identique, ici, puisqu'il s'agit d'une utilisation "savante".

Ces deux sanctions sont les deux seules possibles d'un examen. Contrairement à ce que laisse supposer une confusion fréquente, les autres sanctions ne sont pas le fait d'un examen

- Classer : (distribuer par classe) est le fait d'un concours ou d'une composition
- Sélectionner : faire un choix raisonnable : le fait d'un concours.

Classer et sélectionner ne sont pas nécessairement des opérations inutiles, mais ce n'est pas à l'examen de les effectuer. La conséquence, c'est que, dans un examen, les notes sont inutiles, alors qu'elles sont le fondement du système des concours.

Numerus clausus : c'est une combinaison de sélection (aptitude) et de classement (les meilleurs) dans le but de choisir non en seule fonction de la valeur mais en superposant un classement à un nombre d'acceptés fixé a priori : ce qui différencie cette opération formellement d'un examen. C'est une adaptation d'un volume à une capacité : d'un contenu au contenant. C'est une opération administrative, politique (dans le sens le plus large) et non d'enseignement.

SCORE

Terme anglais signifiant, en français, "résultat d'un match exprimé par le nombre de points qu'a obtenu chaque participant".

Puisque c'est un terme anglais, la recherche de sa signification réelle n'est pas inutile. La traduction la plus générale qu'on puisse en donner est "entaille" sans se soucier ni du lieu, ni de la profondeur, ni du motif. Ainsi "to run a score at a public house" signifie "avoir une ardoise au bistrot". On peut rapprocher du terme anglais score, le terme français "scarifier" qui signifie bien faire une entaille. Faire une entaille (ou une encoche) c'est "cocher". Si "scorer" n'est pas français ce néologisme est inutile puisque "cocher" existe. Ainsi, le premier temps proposé dans la démarche qui aboutit à la notation s'appelera-t-il "établissement du score". Cette opération consiste d'abord à cocher les item reconnus acceptables et à les compter ensuite sans qu'il soit a priori possible d'attribuer une valeur à ces encoches.

On peut penser qu'au bon vieux temps des pubs enfumés et des tables en bois, le cabaretier cochait le compte du client à chaque consommation. La facture devait être finalement différente si le "score" s'exprimait en pintes de bière, en flacons de whiskey ou en verres de gin.

EXAMEN

L'examen, c'est l'action de considérer avec attention et pour se rendre compte. C'est aussi une épreuve, orale ou écrite que subit un candidat pour être admis à tel ou tel grade. Ce mot présente deux sens, différents parce que considérés de deux points de vue : celui des candidats et celui des examinateurs. Le terme "subit" indique que le candidat n'a pas la maîtrise de l'épreuve dans sa conception ni dans le jugement qui en est l'aboutissement. Il n'est pas indispensable de traduire subir par "passer sous les fourches caudines". Le "vae victis" ne semble pas trahir une certaine conception des examens dans laquelle l'épreuve doit être rigoureuse, douloureuse .... peut-être expiatoire. Il n'est pas certain que l'attitude sadique que les candidats reprochent aux examinateurs soit toujours mal fondée. Une certaine complaisance des examinés n'est pas davantage à exclure. Les choses considérées sous cet aspect ne donnent probablement pas une image parfaitement fidèle de la réalité, mais, si cela était, même en partie, ce serait l'épreuve pour l'épreuve. Son but et son intention auraient été oubliés. L'examen serait mauvais et personne ne le nierait s'il était démontré que son esprit était bien celui qui peut être résumé par "drame sado masochiste". Pour dure que soit cette expression, l'explication d'une telle position ne semble pas hors de portée. Mais qu'est-ce donc qu'un bon examen ? Tout dépend de l'idée qu'on s'en fait, de la fonction qu'on lui attribue et de la définition qu'on en retient.

C'est une formalité puisque, en règle, le franchissement d'un degré ne peut être légalement ou réglementairement autorisé qu'à la suite d'épreuves définies. Si ce n'est qu'une formalité, pour être bon, il lui suffit d'être. Il pourrait donc être défini par sa date, sa longueur et son sujet (sa forme). S'il est autre chose, en plus, il doit être défini par un contenu : un fond. Le fond, c'est ce qu'apporte le candidat et que l'examinateur considérera avec attention pour se rendre compte et rendre compte (au moins au jury). Le fond, c'est ce que montre le candidat au cours d'une épreuve, par définition destinée à porter une preuve. Une preuve de quoi ?

A - L'examen est une épreuve d'aptitude

Son but est pronostique. Sa réalisation exige qu'il soit répondu à des questions difficiles : quoi pronostiquer ? Comment pronostiquer ? Quand pronostiquer ? Bien évidemment, avant de se prononcer sur une méthode, il faut que le but soit défini. Dans ce domaine, les réponses sont très variables suivant les cas, mais il est possible de schématiser.

Le but recherché peut être soit de mettre en évidence une aptitude (ou une inaptitude) aux études menant à un métier, soit, des aptitudes au métier lui-même. Cela suppose que le métier soit défini en termes tels qu'une exploration dans un but pronostique à son sujet ne soit pas vaine. Quant aux études, il faudrait encore en savoir si elles représentent une fin en soi ou un moyen de parvenir au métier.

Un exemple célèbre illustre le divorce fréquent entre ces deux objectifs. On dit que Bonaparte ne serait admis dans aucune de nos modernes écoles militaires, son orthographe étant par trop défectueuse. Ainsi, une inaptitude (peut-être réelle, d'ailleurs) aux études militaires entraînerait ipso facto une déclaration d'inaptitude au métier militaire, alors que, même les non bonapartistes, ne refusent à Napoléon ni l'aptitude, ni des capacités certaines (acquises) dans l'art militaire. Dans le même ordre d'idée, il semble curieux, pour le profane, d'apprendre que tel apprenti serrurier ne pourra jamais obtenir de qualification professionnelle digne de ce nom à cause d'une orthographe défectueuse.

S'agit-il d'études supérieures ? C'est, en principe, le baccalauréat qui répond de ce genre d'aptitude. Il est vrai que les études supérieures sont tellement variées qu'on ne peut raisonnablement exiger d'un seul et même examen qu'il discerne le futur chartiste, le futur juriste, le futur médecin ou le futur physicien. Il ne semble donc pas ridicule qu'un examen, à un certain moment, cherche à déceler des aptitudes particulières : ce serait alors une sélection. Entreprendre une sélection suppose qu'on ait répondu à quelques questions :

- Sélection pour les études ou pour le métier ? Admettre une différence pose la question redoutable de l'adaptation des études à leur but.
- Sélection comment et quand ? Bien évidemment, tôt : ce n'est pas au terme de sept années d'études qui ne mènent à rien d'autre qu'on peut sérieusement déclarer que tel candidat est inapte aux études ! Très tôt, quand il faut, implique qu'on est certain que le sujet examiné (jeune par définition) ne changera en rien d'essentiel. Or, H. PIERON,

considéré comme orfèvre en la matière, reconnaît que "la science psychologique n'a malheureusement pas encore abouti à des résultats définitifs". En persistant à ignorer l'ignorance actuelle et avouée des spécialistes, on se demande à quoi ils servent puisque tant d'enseignants se déclarent parfaitement capables de faire un tel pronostic.

Il est vrai, que la méthode souvent employée est assez subtile. Ainsi, en Médecine, un premier cycle de deux ans, est une introduction aux études. La première année du second cycle est une révision des généralités déjà étudiées et l'introduction de nouvelles généralités. La dernière année est consacrée à la récapitulation. Restent deux ou trois ans pour apprendre. Trois ans sont consacrés à "apprendre à apprendre". Ceux qui sont dégoûtés s'en vont. Ceux qui ont persisté, avec de brillants résultats parfois (mais que valent-ils ?) sont déclarés aptes aux études. N'est-il pas souverainement illogique de déclarer aptes à des études de pathologie des sujets qui ont été brillants en physique ? Dans des études réputées plus modestes et qui sont, en tout cas, moins longues, les mêmes défauts sautent aux yeux pour peu qu'on les tienne ouverts.

Une sélection considérée en fonction de son but : établir un pronostic, ne peut être raisonnablement faite qu'après qu'on sache la faire. Si on estime savoir, elle est utile : elle empêche de se fourvoyer un étudiant qui n'a aucune chance raisonnablement prévisible de mener ses études à leur terme et avec succès.

Voltaire, dans Zadig, présente une excellente méthode de sélection. C'est peut-être lui que Beaumarchais persiflait : "il fallait un financier, ce fut un danseur qui l'obtint". Mais Zadig avait beaucoup de chance : il y avait une place et un seul candidat qui présentât la qualité requise par le roi. Si deux candidats avaient été jugés aptes, il y eût eu un "numerus clausus" : il aurait fallu choisir, non sur la seule aptitude mais sur sa grandeur, pour établir une comparaison. Ce qui aurait supposé que fût mesurable l'aptitude. Et M. de Voltaire était bien trop intelligent pour se fourrer dans une aussi méchante affaire.

B - L'examen est une épreuve de capacité

Reste l'aspect "diagnostic" dont on peut affirmer qu'il prend, à un certain moment, une importance capitale. Diagnostic de quoi ? De ce qu'on appelle "connaissances". Quel que soit son niveau, il ne

fait de doute à personne qu'un tel examen se passe dans un cadre scolaire : c'est-à-dire que le candidat doit apporter la preuve qu'il a, dans le domaine cognitif et, parfois, psychomoteur, acquis quelque chose qui, par définition, n'était pas inné. Cette "valeur ajoutée" est le fruit d'un enseignement, fût-il autodidactique, et ceci amène à se poser la question des rapports qui unissent, ou devraient unir, l'enseignement et l'examen. Deux conceptions, a priori, peuvent se défendre : l'examen est <u>en</u> rapport, ou <u>sans</u> rapport avec l'enseignement.

a) <u>Sans rapport</u>. Ce n'est pas une conception paradoxale. On pose n'importe quelle question, unique ou multiple, sur n'importe quel sujet : par exemple "définition d'un véritable travail de romain". Il est de la plus élémentaire honnêteté que le jugement s'établisse en dehors de toute référence à l'éventuel cours, sans quoi, l'épreuve ne serait que de flagornerie. Et si cela était, ce qui peut se défendre, il s'agirait d'une épreuve <u>en</u> rapport avec l'enseignement. Un tel examen est sûrement très difficile à juger : la correction est subjective presque par définition. Mais, tel qu'il est, son intérêt peut être certain, voire évident en tant que test des moyens d'expression et de culture générale, les deux étant d'ailleurs fortement liés. Cependant, faire de ce genre d'examen, un examen d'aptitude est dangereux : l'expression n'est pas tout (il n'est pas nécessaire de tenir les choses pour en discourir).

b) <u>En rapport</u>. C'est ce qui est généralement souhaité. Contrairement aux "examens de dépaysement" précédents, ceux-ci sont liés à l'enseignement : devront l'être par le fond et non par la forme. Répéter servilement ce qu'a dit le professeur ne prouve rien, sinon une bonne mémoire, une aptitude à écrire vite et lisiblement et, peut-être, aussi, l'incapacité, voire l'inaptitude à s'échapper d'une certaine forme pour s'exprimer : celle du texte de référence. Peut-être, encore, dans une certaine mesure, l'intensité du charme subi par l'étudiant et ce n'est pas nécessairement un avantage.

Si enseignement et examen sont en rapport, ils ne peuvent guère l'être que par leurs objectifs communs. Ce qui implique que l'enseignant en ait une claire conscience, sans quoi ses chances de les

atteindre sont médiocres. La moindre des choses qu'on puisse attendre d'un enseignant, c'est qu'il sache pourquoi (dans quel but) il enseigne : connaissances purement factuelles, compréhension, application ... Il est aussi nécessaire que ceux pour qui l'enseignement est fait : les étudiants, le sachent.

Si enseignement et examen sont en rapport, l'activité de l'enseignement et celle des enseignés sont coordonnées puisque, pour l'un comme pour les autres, un but commun est défini. En ajoutant à cela une confusion entre objectif des études (un métier, par exemple ou quelque chose d'autre de définissable) et celui ou ceux de l'enseignement, toutes activités coordonnées, il y a quelques chances pour que le rendement de l'opération s'améliore, que le jugement, dernier acte de l'examen, ait une signification et que les épreuves prouvent quelque chose.

B DONNER

UN SENS AUX RESULTATS

UNE SIGNIFICATION AUX NOTES

" Celuy me semble avoir très bien conceu la force de la coustume, qui premier forgea ce conte, qu'une femme de village, ayant apris de caresser et porter entre ses bras un veau dès l'heure de sa naissance, et continuant tousjours à ce faire, gaigna cela par l'accoustumance, que tout grand boeuf qu'il estoit, elle le portait encore. Car c'est à la vérité une violente et traistresse maistresse d'école que la coustume".

MONTAIGNE.

En appelant " copie " la production que fournit un candidat en réponse à une épreuve,en rappelant qu'une épreuve est une opération organisée aux fins de révéler des preuves,l'analyse d'une copie aboutit à un jugement pouvant prendre deux formes:

- Un résultat ou
- Une note.

Le résultat est une sanction binaire pouvant s'exprimer par " accepté " ou " refusé " et signifiant " suffisant " ou " insuffisant ".

La note est un jugement nuancé précisant le degré d'éloignement du candidat jugé,dans un sens ou dans l'autre,du " seuil de suffisance " généralement appelé " moyenne ".

Le résultat doit être établi mais,selon le cas,il sera considéré comme une fin ou comme une étape.

CHAPITRE II : ETABLISSEMENT DU RESULTAT

CHAPITRE III: DU RESULTAT A LA NOTE

CHAPITRE IV : UNE NOTATION NORMALISEE

CHAPITRE II

ETABLISSEMENT DU RESULTAT

B- DONNER UN SENS AUX RESULTATS

LE RESULTAT 182
UN SECOND PRINCIPE D'ARCHIMEDE 183
COMMENT DETERMINER LE CARACTERE SUFFISANT D'UN CANDIDAT? 185
SIGNIFICATION DE LA MOYENNE 186
LES ITEM ONT-ILS TOUS LA MEME VALEUR ? 187
PRINCIPES D'ETABLISSEMENT DU RESULTAT 189
UNE METHODE:RESULTATS ETABLIS SUR LA REUSSITE (RER) 191
ELEMENTS DE CRITIQUE DE LA METHODE 193
ESSAYONS DE FAIRE AUTREMENT 195
PROCES DE LA FACILITE 200
CORRECTION BINAIRE ET NUANCES 202
QUALITES DE LA METHODE RER 203
APPLICATION D'UN COEFFICIENT NORMATIF 205
AVANTAGES D'UN CRITERE NORMATIF DE REUSSITE 208
FAUT-IL ETRE NORMATIF OU PRAGMATIQUE ? 210
COUTUME ET REALISME 212
CONCLUSION 213

Le " résultat ",c'est la " res ultima ",la dernière chose, la seule qui,finalement,intéresse un candidat. C'est ce dont dépend son sort,son avenir,sa vie. Il s'exprime par une décision irrévocable (accepté ou refusé) de caractère binaire et non ternaire: le"je ne sais pas"est inadmissible. Il peut se symboliser par (+) ou (-) selon le sens de la sanction. Il est l'aboutissement d'un jugement.

Le jugement d'une copie résulte fondamentalement d'une comparaison avec un modèle qui peut être,ou non,préconçu. Une telle comparaison n'est pas une mesure à proprement parler mais la démarche intellectuelle qu'elle nécessite est de la nature de celle qui préside à une mesure. De même que la température n'est pas une grandeur mesurable mais que l'acte consistant à prendre connaissance de la valeur d'une température est,techniquement parlant,une mesure,de même,l'établissement du résultat d'une copie est assimilable à une mesure. Ce sont donc les critères habituels de la qualité d'une mesure qui serviront d'éléments de critique de la méthode.

Deux perspectives se manifestent comme très différentes; chacune est concrétisée par une forme différente du " coefficient C " exprimant la valeur respective de chaque item: " N " pour " NORMATIF ", " R " pour " REUSSITE ". Ce qui fait la différence de perspective,c'est que le coefficient représentatif de la valeur de chaque item est établi à priori (N) ou à posteriori (R). Un point commun,cependant,entre ces deux conceptions:la valeur d'un item est déterminée par son importance. C'est dans la détermination de l'importance qu'il y a divergence. Cette différence de conception n'influe en rien la méthode de calcul de la note mais change considérablement le résultat!

UN SECOND PRINCIPE D'ARCHIMEDE

Un examinateur doit comparer un candidat à un modèle. On attend de lui qu'il réponde aux questions suivantes:

1° - le candidat est-il conforme au modèle?
2° - s'en éloigne-t-il en plus ou en moins?
3° - de quelle grandeur?

On peut proposer une analogie: la " surface " du candidat est-elle égale,supérieure ou inférieure à celle du modèle? Si les surfaces sont " simples " (carré ou rectangle) il suffit de mesurer la longueur des côtés. Si la surface est " compliquée " (un cercle) il suffit de mesurer le rayon.....à condition de connaître la valeur de π . Or,nous ne connaissons pas la valeur de π ce qui interdit de comparer la surface d'un carré et celle d'un cercle,la valeur du candidat et celle du modèle.

Un problème de même nature s'est posé aux Grecs d'il y a 25 siècles:c'était la quadrature du cercle. Il s'agissait de construire un carré ayant la même surface que celle d'un cercle donné. Ce qui revient à comparer,dans des conditions définies,deux surfaces: l'une simple et connue (simple parce que connue),le carré,et l'autre difficile parce qu'inconnue: le cercle.

HIPPOCRATE (de Chio) eut,le premier,semble-t-il,l'idée de décomposer le cercle en petits fragments dont les aires étaient mesurables. Puisque,se dit-il,il s'agit de comparer deux surfaces encore faut il les connaître toutes les deux. En additionnant la valeur des aires élémentaires il espérait trouver la grandeur cherchée. Mais,toujours, il persistait un résidu non mesurable: le résultat n'était qu'approximatif. Cependant,l'évaluation de la surface,impossible à obtenir par une seule multiplication,faute de connaître la valeur de π,se faisait par addition. C'est bien ainsi que sont " mesurés " les résultats d'une épreuve.

Quelques années plus tard,ANTIPHON puis ARCHIMEDE,eurent l'idée d'aborder le problème par la circonférence. Puisque deux cercles égaux ont même circonférence,il vaut mieux s'attaquer à des dimensions linéaires. Linéaires mais non rectilignes. Ceci n'a pas beaucoup d'importance,pensa ANTIPHON. Prenez un triangle équilatéral inscrit dans un cercle: son périmètre est différent et plus petit que celui de la circonférence. Doublez le nombre de côtés,et encore....l'hexagone,puis le dodécagone obtenus ont toujours un périmètre plus petit mais d'autant moins différent que le nombre de côtés est plus élevé. Soit,dit ARCHIMEDE qui passa jusqu'au polygone à 384 côtés (3.2^7). Mais si je sais que la valeur mesurée est toujours inférieure,je ne sais pas de combien. Et il bâtit le polygone ex-inscrit: celui-là exagère toujours,dit-il, mais d'autant moins que le nombre de ses côtés est grand. Ainsi,en comparant un périmètre un peu trop grand et un périmètre un peu trop petit, je sais que la circonférence présente une valeur " coincée ",non mesurable exactement,certes,mais je sais dans les limites de quelle erreur. Il aboutit d'ailleurs à une assez bonne approximation de π dont la valeur ne fut connue à la 30è décimale qu'à la fin du XVIè siècle (ROMAIN, METIUS,VAN CEULEN).

De cette histoire,on peut retenir quelques leçons:

1° - Il est vain de chercher l'exactitude là où elle n'est pas: même au XXè siècle,on ne peut mesurer exactement la surface d'un cercle.

2° - On fait moins d'erreur en comparant deux circonférences mesurées que deux cercles " évalués à l'oeil ".

3° - Plus un polygone possède un nombre important de côtés et moins on fait d'erreur en le comparant à la circonférence correspondante.

4° - Faute de connaître un facteur de proportionnalité simple permettant de déduire une dimension complexe à partir d'une dimension simple et facile à mesurer,on en est réduit aux additions.

5° - Plus une dimension est facile à mesurer et plus on peut se fier aux résultats de la mesure: c'est ce qu'avait compris ARCHIMEDE en "confondant " un arc et sa corde.

Au total,on peut diminuer les erreurs de la correction en multipliant les mesures et,dans le cas précis,en utilisant un grand nombre (au moins 30) d'unités comptables (item): c'est le second principe d'ARCHIMEDE.

COMMENT DETERMINER LE CARACTERE SUFFISANT D'UN CANDIDAT?

Les critères retenus pour définir le modèle (caractère suffisant) peuvent être externes ou internes à la population dont fait partie le candidat. Les critères internes peuvent être rigoureux,très vagues ou fondamentalement faux. A ce sujet,deux mythes semblent particulièrement nocifs: celui de la"moyenne"et celui de la "répartition gaussienne".

L'examinateur juge le candidat. Les critères peuvent être propres au sujet (l'examinateur): ils sont subjectifs. Seuls peuvent être objectifs ceux qui dépendent du candidat. Ainsi,le seul fait que les critères soient externes ne suffit pas à garantir leur objectivité. Eliminés les mythes auxquels beaucoup d'examinateurs vouent un véritable culte,il ne reste guère que deux solutions:

a) juger un candidat par rapport à ce qu'il doit être au minimum. Ceci suppose qu'un tel modèle soit défini sur des critères réalistes,fondés sur la situation réelle: par exemple qu'on sache (et qu'on décrive pour montrer qu'on sait) ce qu'on attend de lui. Condition absolument nécessaire à la sanction qui consiste à lui faire savoir qu'on est satisfait ou déçu.

b) juger un candidat par rapport à la moyenne de la population à laquelle il appartient. Il revient ainsi à la population même de définir le candidat suffisant,le modèle minimal dont la valeur représentera le seuil de suffisance qu'on pourra ensuite,et sion le juge utile, exprimer par 10/20. Encore faut-il connaître la valeur de cette moyenne. Si on choisit comme référence "le candidat moyen"il faut le définir rigoureusement.

Si le critère interne est retenu,il sera défini en termes de réussite. Le candidat modèle (il faut comprendre le "candidat de référence") sera défini par une grandeur mesurable,exprimée en nombre d'item réussis,il sera appelé "candidat moyen" sans craindre que ce mot ne cache quelque vague notion subjective,affective ou sentimentale. Dans tous les cas,sa valeur sera exprimée par zéro.

Celà implique que l'image de l'examen soit une pesée faite sur une balance "à zéro". Sur un plateau le "candidat moyen",la tare. Sur l'autre,le candidat réel: s'il"fait le poids" l'aiguille déviera du "bon côté": (+).Le candidat est reçu. Sinon,la déviation se fait vers (-), le candidat est récusé: il ne"fait pas le poids". Peu importe,au fond, de savoir de combien il dépasse la tare ou ce qui lui manque pour en atteindre la valeur.

Pour que l'opération soit logique,il faut que réponse soit donnée à deux questions:

- Quelle est la signification de la moyenne?
- Les item ont-ils tous la même valeur?

SIGNIFICATION DE LA MOYENNE

On pose à un certain nombre de candidats deux questions. Les uns fournissent deux réponses justes: ils sont reçus. Les autres deux réponses fausses: ils sont refusés. Certains,une seule réponse juste: ils sont "moyens". Aux premiers,on attribuera + 2,aux seconds: -2; aux troisièmes: 0.

Question:Les candidats "moyens" doivent-ils être refusés ou acceptés?

1ère réponse:

"Je ne sais pas parce que ces candidats s'éloignent autant des reçus certains que des certainement refusés". Comme un examen doit avoir un résultat,positif ou négatif et non interrogatif,deux solutions se présentent:

- on tire à pile ou face
- on s'en remet à un critère extérieur à l'épreuve (par exemple,le résultat d'une autre épreuve). Critère jugé déterminant par conséquent.

2ème réponse:

"Oui,parce que ces candidats en savent plus que ceux qui sont certainement refusés". Mais alors,quelle serait votre réponse pour une réussite de 2 sur 4? Logiquement il faudrait même accepter ceux qui ont réussi 1 sur 4 puisqu'ils savaient quelque chose.

3ème réponse:

"Non,parce que recevoir à 1 sur 2 ce serait aussi inacceptable que de recevoir à 6 sur 20: ils savent quelque chose mais pas assez". Celà signifie que pour être reçu il faut en savoir assez et pour être refusé en ignorer trop. Mais combien dans le cas précis : réussite à une question sur deux?

4ème réponse:

"Celà dépend. Votre question est mal posée,insidieuse,perfide. Il y a dans votre combinaison,une seule manière d'être certainement reçu: réussite de 2 sur 2. Et une seule manière d'être certainement refusé: échec de 2 sur 2. Mais il y a deux manières d'être "moyen": réussir à l'une ou l'autre des questions. Je donnerai la préférence,éventuellement,à telle ou telle des deux catégories de "moyens" lorsque je connaîtrai l'importance relative des deux questions".

LES ITEM ONT-ILS TOUS LA MÊME VALEUR?

Tout dépend du critère de valeur. On peut les différencier entre eux par la longueur de la question et de la réponse: on aurait alors les "grandes" et les "petites" questions,et l'unité de valeur pourrait être le nombre de mots ou le temps passé à lire la question ou à y répondre. Ce qui est évidemment et implicitement entendu dans le terme de "valeur" d'un item est en rapport avec l'enseignement par nature. C'est un critère d'importance. On peut supposer que,si la réponse doit être longue,c'est qu'elle doit être composée d'éléments plus nombreux qu'une réponse courte. Ces éléments,pris en compte,sont précisément les item. Il est évident qu'une "grande question" a plus d'importance qu'une "petite question" par le nombre d'item quelle contient mais, à supposer que deux questions appellent en réponse le même nombre d'item, la valeur de chacune des questions dépendra finalement,non du nombre des item (c'est le même) mais de la valeur de chacun d'eux. Chaque réponse étant considérée comme un ensemble d'item,la valeur de l'ensemble dépend non seulement du nombre d'éléments mais encore de la valeur de chacun d'eux. Si on estime que le critère de valeur,c'est l'importance, ceci suppose qu'on puisse attribuer une valeur à l'importance.

A priori,trois solutions sont possibles:

1°) On décompose la réponse à faire en éléments,on dénombre ensuite ces éléments et le nombre obtenu détermine la valeur. Deux écueils s'attachent à cette méthode. D'une part la tentation est grande d'attribuer à priori une valeur exprimée en nombre de points à la question puis, secondairement,décomposer la question en autant d'item qu'il y a de points prévus. La valeur de l'item est ainsi circonstancielle. La valeur de telle notion dépendra finalement de son contexe (question longue ou brève).

D'autre part,et au moins,dans le cadre d'un ensemble d'item représentant une "question",ne prendre que le nombre en considération revient à dire que tous les item de l'ensemble (une question ou une épreuve) ont la même valeur. Attribuer la même valeur à chacun des item constituant un ensemble,relève d'une opération arbitraire et fausse si on ne sait pas qu'ils ont réellement la même valeur.

De toutes façons,les item seront dotés d'une valeur attribuée de façon arbitraire,illogique,aléatoire.

2°) On attribue une valeur à chaque item en fonction de son importance. Deux cas se présentent: ou on connait cette importance et la solution est bonne,ou on ne la connait pas et,malgré le travail supplémentaire fourni,le résultat est aussi arbitraire,illogique et aléatoire que celui qui était fourni par la solution précédente. Le seul critère de fiabilité de ce coefficient de valeur arbitrairement attribué,c'est que tout arbitre d'un moment à l'autre,et que tous les arbitres au même moment lui attribuent la même valeur. A condition,bien entendu,que les juges soient compétents en la matière et que leur verdict ne soit pas systématiquement 0,0.5 ou 1.

3°) On attribue à chaque item une valeur dépendant directement de leur réussite par les candidats. Ainsi,la valeur retenue sera, pour chacun,le coefficient de réussite.

En supposant qu'on sache "pourquoi" les item sont d'inégales valeurs,on sait très bien "comment" l'exprimer: en leur affectant un coefficient,qu'il est possible de moduler: 1 pour la plus grande valeur possible,0 pour l'item sans valeur,0.9 ou 0.8 pour des item relativement importants ou,plus finement encore,0.79 ou 0.78...etc. L'important c'est de pouvoir justifier la valeur choisie:0.5 doit bien exprimer une valeur supérieure à 0.4 et inférieure à 0.6. La précision des résultats dépend finalement de la précision de ce choix; une indécision telle que: "0.5 est peu différent de 0.5 ± 0.1" entraîne une imprécision de 20p.100 en plus ou en moins. Celà revient à dire qu'il n' y a aucune différence nette entre quatorze sur vingt et six sur vingt.

PRINCIPES D'ETABLISSEMENT DU RESULTAT

1) PRINCIPES GENERAUX.

La "moyenne" (référence) est zéro. Chaque item réussi sera marqué (coché)(+),chaque item manqué sera marqué (-). Le résultat final sera exprimé essentiellement par un signe et accessoirement par une valeur. Le modèle peut être représenté par un polygone (éventuellement irrégulier) dont le périmètre représenterait la somme minimale des valeurs à obtenir. La valeur de chaque candidat est le résultat d'une addition: c'est la somme (algébrique) de ses réussites et de ses échecs. Se trouve donc appliqué le second principe d'ARCHIMEDE par lequel la comparaison de deux surfaces est remplacée par la comparaison de deux périmètres qu'il n'estpas absurde d'espérer connaître par une suite d'additions.

a) La moyenne (seuil de suffisance) est zéro.
b) Les item réussis sont positifs,les item manqués négatifs.
c) Le résultat s'exprime par le sens de la somme ($\pm$).

2) RESULTAT ET VALEUR DES ITEM.

Dans l'absolu des examens.et non dans le relatif d'une épreuve,les item n'ont pas tous la même valeur. Dans une épreuve ou une partie d'épreuve,l'égalité de valeurs des différents item est un cas particulier parfaitement concevable et qui ne change rien aux principes. Si tous les item d'une même épreuve ont même valeur,cette valeur peut être très grande (0.9),très petite (0.1) ou intermédiaire. Ainsi,quelle que soit la constitution de l'épreuve,ces deux principes seront appliqués:

PREMIER PRINCIPE: La réussite ne doit pas dépendre uniquement du nombre d'item réussis.

DEUXIEME PRINCIPE: La réussite dans un item peu important ne peut pas compenser l'échec dans un item important.

Attribuer à un item une valeur 0.9 (quelle soit la méthode d'attribution) celà signifie qu'on le juge important. Puisqu'on exprime cette importance par un nombre,celà veut dire qu'on admet un échec sur 10 item d'une telle valeur. Alors que la valeur 0.9 exige une réussite de 9/10,une valeur de 0.4 n'exige qu'une réussite de 4/10 dans une épreuve constituée d'item de valeur 0.4.

La réussite à un item important rapporte peu,l'échec coûte cher. L'échec dans un item peu important coûte peu,sa réussite rapporte plus qu'elle ne coûte.

Il est normal de réussir ce qui est important puisque l'importance est la marque d'une exigence: c'est pour celà que le "rapport" est d'autant moins grand que la valeur est plus grande. Inversement,la pénalisation est d'autant plus grande,en cas d'échec,que la valeur attribuée à l'importance était plus grande. Quant à l'obstacle mis dans la tentative de compensation de manques dans ce qui est important par des réussites dans ce qui ne l'est pas,c'est simplement logique.

SCORE	ITEM N° 1	ITEM N° 2		ITEM N° n
Candidat NO 1	I	0	...	I
Candidat N° 2	0	I	...	0
	...	...	...	...
Candidat N° n	0	I	...	I
Nombre de candidats ayant réussi l'item	R'_{i_1}	R'_{i_2}		R'_{i_n}
Coefficient de réussite de l'item	R_{i_1}	R_{i_2}		R_{i_n}
Carré du coefficient	$R^2_{i_1}$ +	$R^2_{i_2}$ +	.....	$R^2_{i_n}$ + ... = ΣR^2

SCORES PONDERES INDIVIDUELS (S.P.I.)

Candidat N° 1: $(I \times R_{i_1} + 0 \times R_{i_2} + I \times R_{i_3} + ? \times R_{i_n}) - \Sigma R^2$

Candidat N° 2: $(\quad R_{i_2} + \quad ? \times R_{i_n}) - \Sigma R^2$

Candidat N° n: $(\quad R_{i_2} + \quad R_{i_3} + ? \times R_{i_n}) - \Sigma R^2$

Le résultat est fourni par le SIGNE de la différence: $\Sigma R_c - \Sigma R^2$

UNE METHODE: RESULTATS ETABLIS SUR LA REUSSITE.

Lorsque G.BARRES,en septembre 1971,publia une nouvelle méthode d'établissement des résultats,on ne peut pas dire qu'il a déchaîné l'enthousiasme des foules. Il y eut,à vrai dire,très peu d'opposition et même très peu d'objections. Peu furent convaincus. Ce qui fut le plus clair,dans l'ensemble,ce fut l'absence totale de réactions! Ceci suscite plusieurs réflexions. Il est évident que la plupart de ceux qui eurent ce texte sous les yeux,ne le lirent pas,ce qui montre l'intérêt attaché, par ceux qui en ont la charge,pour ces sortes de questions. Parmi les rares lecteurs,la plupart furent rebutés par des formules,pourtant autrement simples que celles que maints d'entre eux manient à la journée: on veut bien exprimer des résultats par des grandeurs mais on ne veut pas compter. On veut bien appliquer des recettes mais on ne veut pas chercher de lois. On veut bien agir mais pas se poser de questions sur le sens de l'action,peut être parce que chacun est plus ou moins conscient que s'interroger sur cette sorte de sujets risquerait d'emmener très loin,trop loin et de déranger la routine. Enfin quelques uns,extrêmement rares,qui déclarèrent avoir compris,objectèrent que:

1) celà était bien compliqué.

2) celà était sans nuances.

3) Il était choquant de prendre en considération le succès d'une question pour lui attribuer une valeur proportionnelle. Sans que personne,semble-t-il n'ait osé le proclamer,celà sentait sa démagogie.

En admettant que cette méthode soit compliquée,sans nuances et démagogique,et il ne semble pas qu'elle soit rien de tout celà,il serait nécessaire de la comparer avec les prétendues méthodes couramment employées quand il ne s'agit pas d'absence totale de méthode. De quoi s'agit-il?

1) Sur une grille présentant:
   . Une entrée "candidats": les lignes
   . Une entrée "item": les colonnes

donc sur une grille présentant un nombre de cases qui est le produit Qe.C du nombre de candidats (C) par le nombre d'item dont est constituée l'épreuve (Qe) on coche les réussites. Cette grille présente donc le SCORE.

2) Pour chacun des item (1,2...n) on compte le nombre de réussites (R'_1,R'_2,...R'n). Un item ne connait que deux sanctions: accepté ou refusé.

3) En divisant le nombre de réussites obtenues dans chacun des item (R'_1,R'_2,...R'n) par le nombre de candidats (C) on obtient le coefficient R de réussite de chacun d'eux: R_1,R_2,...Rn selon la formule:

$$R = \frac{R'}{C}$$

4) On élève chacun de ces coefficients ($R_1, R_2, \ldots R_n$) au carré ce qui fournit $R_1^2, R_2^2, \ldots R_n^2$

5) On établit le résultat de chaque candidat sur la démarche théorique suivante:

. L'item N°1, supposé réussi, vaudra + $R_1(1-R_1)$
. L'item N°2, supposé manqué, vaudra - $R_2(R_2) = -R_2^2$

Le résultat est la somme algébrique des réussites (+) et des échecs (-). Un résultat positif désigne le candidat reçu, un résultat négatif récuse le candidat.

PRATIQUEMENT:

1) On établit le score.
2) On détermine le coefficient de réussite de chaque item
3) On fait la somme des carrés de chaque coefficient (forme: ΣR^2) ce qui détermine le plus grand négatif possible et qui représente en valeur absolue le score pondéré minimum à réaliser pour obtenir zéro, donc être reçu: c'est le handicap.
5) On attribue systématiquement $-\Sigma R^2$ à chaque candidat.
6) On fait la somme, pour chaque candidat, des coefficients de réussite de chacun des item qu'il a effectivement réussis.
7) On retranche de la valeur attribuée à chaque candidat (forme: ΣRc) le handicap (forme: ΣR^2) selon la formule:

$$\Sigma Rc - \Sigma R^2$$

ΣRc représentant la somme des réussites du candidat "C"
ΣR^2 représentant la somme des carrés de tous les coefficients de réussite de l'épreuve.

Trois cas se présentent:

. $\Sigma Rc - \Sigma R^2 > 0$ Candidat RECU.
. $\Sigma Rc - \Sigma R^2 < 0$ Candidat RECUSE.
. $\Sigma Rc - \Sigma R^2 = 0$ Candidat douteux, en fait RECU.

mais il sera nécessaire de trancher le point délicat: zéro est-il positif ou négatif.

Pourquoi SIMPLIFIER ?

. Le positif est de la forme $R(1-R)$ soit $R-R^2$ ainsi la valeur attribuée à une réussite est diminuée de R^2.
. Le négatif vaut $-R^2$: ainsi le négatif peut-il se représenter par $0-R_2$ où zéro représenterait l'absence de réussite.
. Que l'item soit réussi ou non, le candidat se verra attribuer $-R^2$
. Si l'item est manqué, le candidat marque $-R^2$.
. Si l'item est réussi, le candidat marque $-R^2+R$.

Ainsi, le résultat est la somme de deux termes: l'un identique pour TOUS les candidats d'une même épreuve et NEGATIF; l'autre, PERSONNEL au candidat, POSITIF.

ELEMENTS DE CRITIQUE DE LA METHODE PROPOSEE

I) LA METHODE EST COMPLIQUEE.

Il faut,à ce sujet,distinguer "compliqué" et "inhabituel". Certes,il est plus facile d'attribuer une note "à l'estime" et de faire de ce sentiment le motif de la sanction que de motiver la sanction sur des critères. Il est plus facile d'exprimer des impressions par des chiffres que d'utiliser une démarche logique pour aboutir à une grandeur rationnellement établie. On a parfaitement le droit de soutenir que la raison n'a rien à faire dans l'établissement d'un résultat. Il serait alors nécessaire de proclamer que,pour être reçu,il faut plaire à ses juges et fournir comme sanction "plaisant" ou "déplaisant". Si on introduit dans la démarche des notions telles que "suffisance","comparaison avec un modèle","valeurs relatives","grandeur de la réussite",la démarche doit être rationelle,faute de quoi elle serait ridicule. Une expression chiffrée représentant une grandeur doit être l'aboutissement d'un raisonnement obéissant à un certain nombre de règles de logique.

Compte tenu des principes appliqués:

- cotation binaire de chaque item et d'un nombre suffisant d'item,
- la réussite ne doit pas dépendre uniquement du nombre d'item réussis,
- la réussite d'un item peu important ne doit pas compenser l'échec dans un item important,

cette méthode ne peut être rejetée comme trop compliquée que si d'autres,plus simples,respectent les mêmes principes. De telles autres méthodes,parfaitement concevables,seront examinées sous le titre "essayons de faire autrement".

II) LA METHODE EST DEMAGOGIQUE.

En effet,le critère d'importance d'un item,c'est son coefficient de réussite. Plus ce coefficient est élevé,et plus l'item s'est avéré facile. Il est parfaitement légitime de considérer que ce qui donne de la valeur à un item,dans cette méthode,c'est sa facilité. On pourrait,sans rien changer,appeler "coefficient de facilité" ce qui est pris comme critère. Ceci impose de faire le procès de la facilité.

Cependant,il ne semble pas inutile de souligner que c'est l'examinateur qui juge du caractère acceptable ou non d'une réponse et que ce n'est pas la réponse faite le plus souvent qui est déclarée "bonne". On peut espérer que la bonne réponse soit la plus fréquente mais ce n'est pas par voie de référendum qu'est établie la vérité.

III) LA METHODE MANQUE DE NUANCES.

Cette méthode a pour seule ambition d'établir un résultat qui s'exprime par une sanction binaire: accepté ou refusé (récusé). Le jugement manque de nuances,c'est vrai et c'est voulu puisqu'il est considéré que la seule question qui se pose est:"ce candidat est-il suffisant?".Les seules réponses attendues sont oui ou non. Même "je ne sais pas" est exclu.

Le jugement est dépourvu de toutes nuances.Mais que doit-on nuancer dans un jugement?. Le verdict ou les attendus?. Si ce sont les motifs du jugement qui doivent être nuancés,l'application de cette méthode introduit des nuances à chaque pas de sa démarche. A condition de ne pas confondre nuance et ornement mais de prendre la nuance pour ce qu'elle est: une précision.

Il n'est pas évident,à priori,qu'un jugement établi sur une démarche primaire de "tout ou rien": un item étant accepté ou refusé, il n'est pas évident,donc,qu'un tel jugement soit nuancé (précis). Cette question sera examinée,elle aussi.

ESSAYONS DE FAIRE AUTREMENT

Afin de ne pas confondre deux sortes de critiques, le coefficient de valeur attribué à chaque item sera simplement désigné par "C". Ainsi seront distingués le mode de calcul et la nature du critère de valeur retenu qui, lui même, ne peut être que choisi par l'examinateur : coefficient normatif (N) ou déterminé par la réussite (R). Même un coefficient purement aléatoire, sorti, par exemple, d'une table de hasard, serait normatif puisque c'est l'examinateur qui aurait choisi le mode de détermination. Le fait qu'il délègue son pouvoir de décision au hasard (ou à l'habitude) ne change rien au fait essentiel : il a fait lui-même son choix, fondamentalement.

La solution à laquelle tout le monde pense : coter (+C) ce qui est réussi et (-C) ce qui est manqué ne sera pas retenue car, à l'évidence, et dans le cas d'une épreuve constituée d'item ayant tous la même valeur (épreuve homogène), le résultat sera toujours satisfaisant pour 50 % de réussite. Que les item soient proclamés très importants(C = 0,9) ou très peu importants (C = 0,1), il suffit d'en réussir la moitié, alors que dans un cas on déclarait accorder le droit à 1 échec pour 10 item et, dans l'autre cas, on se satisfaisait d'une réussite pour 10 item. Une telle solution serait ILLOGIQUE car non en conformité avec ce qui est annoncé ; elle contreviendrait en outre au 1er PRINCIPE puisque la réussite dépendrait uniquement du nombre d'item réussis. Cette solution est à rejeter.

Solution I

En application du second principe, l'importance étant représentée par la valeur du coefficient "C"

. un item manqué sera coté - C
. un item réussi sera côté + (1 - C) (Voir tableau des 3 solutions p.198).

Ainsi, un item très important (C = 1)
. réussi, rapporte 1 (1 - 1) -------- 0
. manqué, coûte - 1 ---------- - 1

Un item très peu important (C = 0,1)
. réussi, rapporte (1 - 0,1)---- + 0.9
. manqué, coûte - 0,1 ---------- - 0.1

Application - Soit une épreuve de trois item : a, b, c, de valeurs respectives : a = 0,6 ; b = 0,4 ; c = 0,2. Trois candidats ; A, B, C, ont réussi respectivement :

A : ab
B : bc
C : ac

Chacun en a réussi deux, mais pas les mêmes. Qui est reçu, avec quel score ?

Score de A : (1 - 0,6) + (1 - 0,4) - 0,2
0,4 + 0,6 - 0,2 = + 0,8 RECU

Score de B : (1 - 0,4) + (1 - 0,2) - 0,6
0,6 + 0,8 - 0,6 = + 0,8 RECU

Score de C : (1 - 0,6) + (1 - 0,2) - 0,4
0,4 + 0,8 - 0,4 = + 0,8 RECU

Tous les candidats ayant réussi deux item sur trois (même score), quelle que soit leur importance, sont reçus avec le même résultat. On peut vérifier que, quelles que soient les valeurs du coefficient (a, b, c) la somme est constante.
Score de A : (1 - a) + (1 - b) - c = 2 - (a + b + c)
Score de B : (1 - b) + (1 - c) - a = 2 - (a + b + c)
Score de C : (1 - a) + (1 - c) - b = 2 - (a + b + c)

Conclusion en opposition avec le premier principe - Cette solution est rejetée : la connaissance d'une question peu importante compense l'ignorance d'une question plus importante.

Solution II

Elle cherchera à éviter l'écueil de la première qui contrevenait au premier principe. Pourquoi ? Les coefficients étaient :
pour a : 0,6
pour b : 0,4
pour c : 0,2
En a : l'échec coûtait - 0,6 ; le succès rapportait (1-0,6) : 0,4
En b : l'échec coûtait - 0,4 ; le succès rapportait (1-0,4) : 0,6
En c : l'échec coûtait - 0,2 ; le succès rapportait (1-0,2) : 0,8

L'écart entre le coût de l'échec et le rapport de la réussite était constant : 1 et, finalement, malgré les apparences tous les item étaient cotés de la même manière. Le défaut à corriger est donc la constance de l'écart. La solution est d'appliquer le coefficient non aux scores (pour les pondérer) mais aux écarts entre coût et rapport (cf tableau des 3 solutions). Pour "a", par exemple, l'écart sera non plus 1 comme dans la solution précédente, mais 0,6. La moyenne étant 0, la réussite rapportera $\frac{+0,6}{2} = +0,3$ et l'échec coûtera $\frac{-0,6}{2} = -0,3$

Que deviennent nos candidats précédents ?
avec : a : $\pm$ 0,3
b : $\pm$ 0,2
c : $\pm$ 0,1

Score de A (ab réussis) : + 0,3 + 0,2 - 0,1 = + 0,4 Reçu
Score de B (bc réussis) : + 0,2 + 0,1 - 0,3 = 0 Reçu
Score de C (ac réussis) : + 0,3 + 0,1 - 0,2 = + 0,2 Reçu

Tout le monde est reçu, certes, mais chacun avec un score pondéré différent, fidèle reflet de la valeur des item réussis. On ne contrevient plus semble-t-il au premier principe.
Conclusion : Malgré les apparences liées à l'exemple choisi, il est clair que cette solution contrevient au premier principe. A supposer que tous les item de l'épreuve aient la même valeur : 0,9 ou 0,1 ; ce qui ferait ou non recevoir les candidats, c'est, finalement, le nombre d'item réussis qu'ils soient importants ou non. Il suffit, dans une épreuve homogène, de réussir 50 p. 100 des item. Or, le premier principe dit que la réussite ne doit pas dépendre du seul nombre d'item réussis. La solution envisagée (II) cotait la réussite à

+ 0.9 et l'échec à - 0.9: bénéfices et pertes sont de même grandeur. Pour un item peu important,la cote aurait été ± 0.1. Dans une épreuve composée uniquement d'item à 0.1,il suffit d'en réussir exactement 50.p.100 pour être reçu. Il en serait de même pour une épreuve uniquement composée d'item à 0.9 (importants). Cette solution,quoique de formulation plus complexe,est l'équivalent de celle à laquelle tout le monde pense et comme elle,elle est à rejeter.

SOLUTION III

Chacune des deux solutions précédentes avait échoué sur le premier principe.

La solution I accordait la même amplitude à l'écart entre la valeur attribuée au coût de l'échec et celle qu'on attribuait au rapport de la réussite. L'amplitude du phénomène n'était pas affectée par le coefficient: le "gain" était constant. Ce qui était modifié, c'est la position du zéro (la moyenne). Le négatif valant (-C),le positif + (1 - C),l'amplitude était donnée par la somme des valeurs absolues des rapports et des pertes: (1-C)+C=1 (gain constant). Dans le sens des conventions habituelles,pour un "C" croissant le négatif se décalait vers la gauche (il augmentait) tandis que le positif se décalait de la même valeur et dans le même sens (cf.le tableau des 3 solutions). La solution I avait pourtant l'avantage de pénaliser l'échec d'autant plus lourdement que l'importance était plus grande mais la contravention au premier principe empêchait que cette intention soit suivie des conséquences souhaitées.

La solution II modifiait le gain,puisque le coefficient était appliqué à l'amplitude de l'écart "valeur absolue de l'échec-valeur de la réussite". Outre le fait que la moyenne se confondait encore avec la médiane,faute de réglage du zéro,cette solution contrevenait évidemment au premier principe dans le cas d'une épreuve homogène. Elle présentait l'avantage particulier de n'accorder qu'une faible valeur aux réussites dans les item peu importants, contrairement à la première. (cf.tableau des 3 solutions).

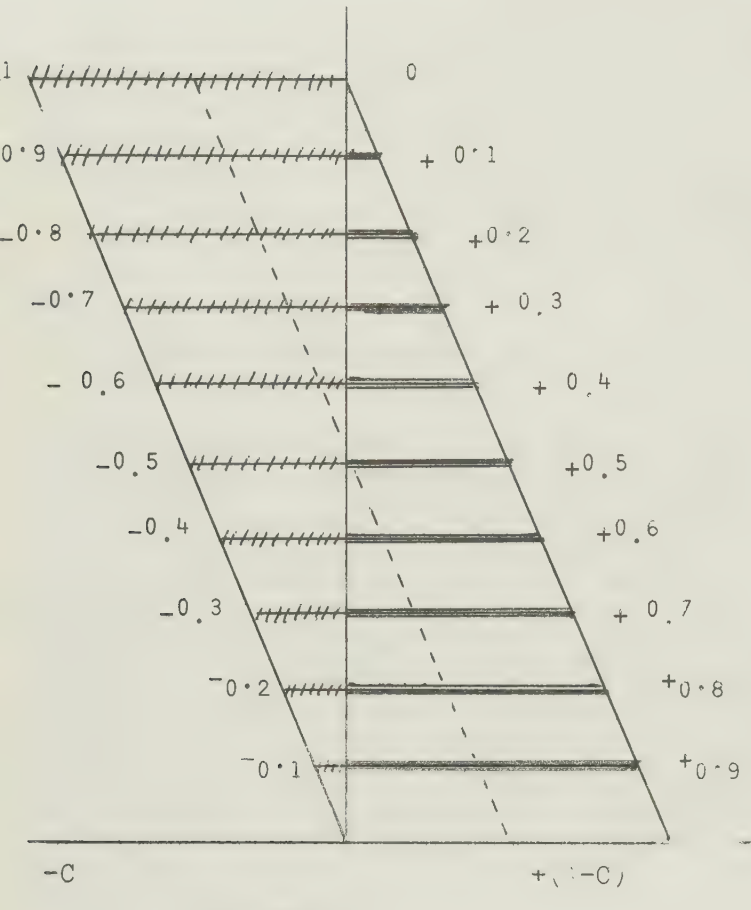
La troisième solution consiste à combiner les avantages des deux premières: décalage du zéro et modification du GAIN. Pour celà il suffit d'appliquer le coefficient là où il a été appliqué dans chacune des deux précédentes solutions, ce qui amène, en plus, la disparition des inconvénients de chacune d'elles.

La première solution définissait:
- . le négatif par (-C)
- . le positif par + (1-C)

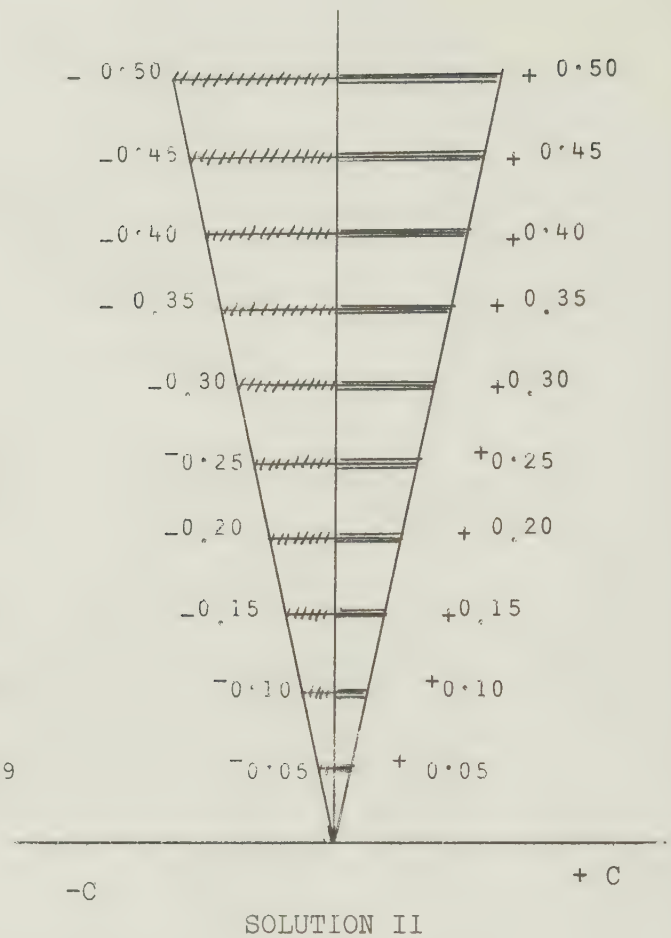
La seconde solution définissait:
- . le négatif par $-\frac{C}{2}$
- . le positif par $+\frac{C}{2}$
- . en affectant "C" à l'amplitude

La troisième solution, produit des deux autres,définit:
- . le négatif par (-C)x C ---------- $-C^2$
- . le positif par (1-C)x C ---------- $+C(1-C)$
- . l'amplitude vaut: $C(1-C) + C(-C)$
  soit: $C-C^2+C^2$
  soit: C

TABLEAU DES 3 SOLUTIONS



SOLUTION I

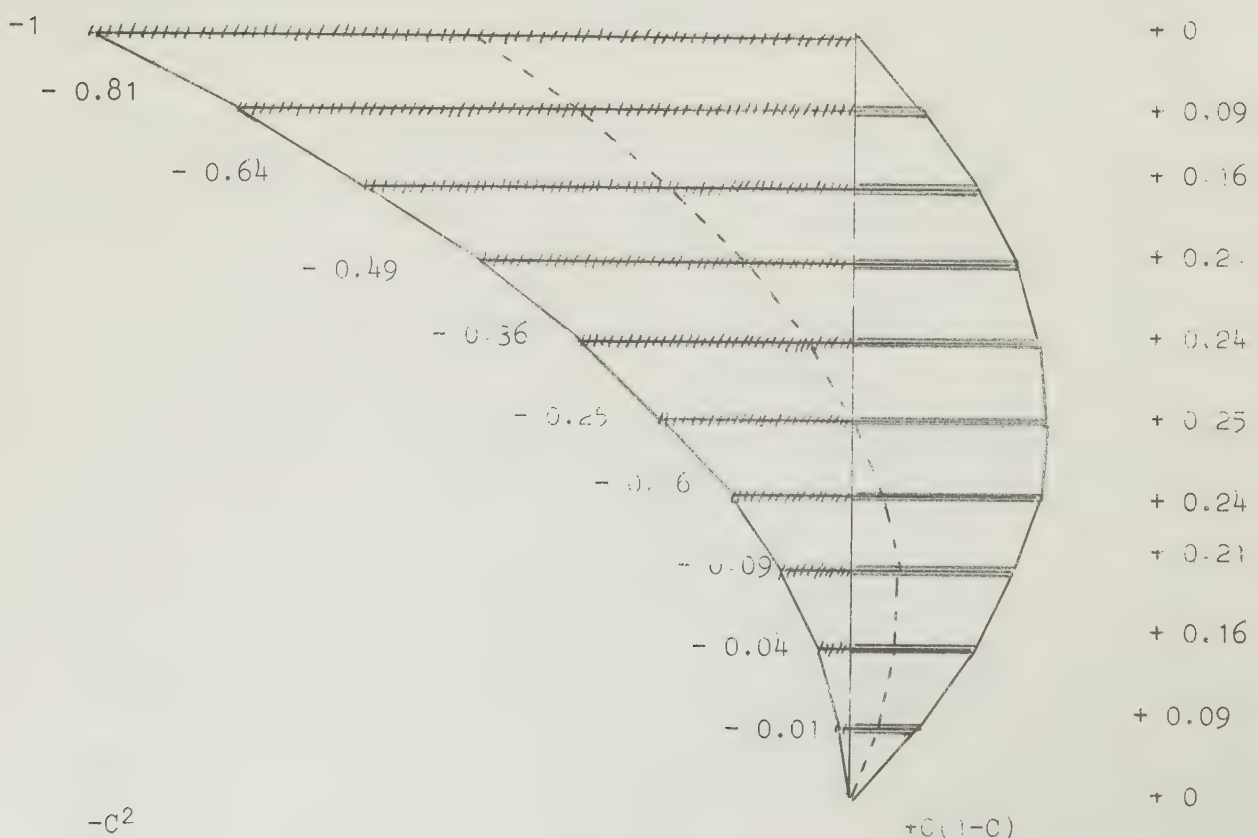


SOLUTION II

//////// Négatif

.....Médiane

SOLUTION III



Que deviennent nos trois candidats?
Score de A: 0.6 - 0.36 + 0.4 -0.16 - 0.04 = + 0.44 Reçu
Score de B: 0.4 - 0.16 + 0.2 - 0.04 - 0.36 = + 0.04 Reçu
Score de C: 0.6 - 0.36 + 0.2 - 0.04 - 0.16 = + 0.24 Reçu

Cette solution obéit aux deux principes: elle est à retenir. De plus,(cf.tableau des 3 solutions) médiane et moyenne ne sont pas confondues par définition. La constatation d'une confusion deviendrait significative puisque cette méthode n'implique pas une telle confusion.

Soit à considérer les sanctions possibles d'item de valeurs différentes: on détermine pour chacun d'eux le minimum: $-C^2$, le maximum: C (1-C), l'étendue de l'écart: C.

Valeur	Minimum	Maximum	Ecart
1	- 1	0	1
0.9	- 0.81	+ 0.09	0.9
0.8	- 0.64	+ 0.16	0.8
0.7	- 0.49	+ 0.21	0.7
0.6	- 0.36	+ 0.24	0.6
0.5	- 0.25	+ 0.25	0.5
0.4	- 0.16	+ 0.24	0.4
0.3	- 0.09	+ 0.21	0.3
0.2	- 0.04	+ 0.16	0.2
0.1	- 0.01	+ 0.09	0.1
0	0	0	0

Un candidat qui aurait raté un item (0.9) aurait - 0.81. Pour compenser cet échec par des réussites à des items (0.1),il devrait en réussir $\frac{0,81}{0,09} = 9$. Le nombre de réussites n'est pas le seul critère (premier principe). Rater un item (0.9) coûte - 0.81, rater un item (0.1) ne coûte que - (0.01) : un item important manqué coûte beaucoup plus cher qu'un item peu important (2ème principe). Ici le rapport est de 81.

PROCES DE LA FACILITE

LA FACILITE EST-ELLE MESURABLE ?

Très facilement. Un item réussi par 90p.100 des candidats est certainement plus facile qu'un item qui serait réussi par 10p.100. L'expression de la facilité d'un item peut être sa"fraction de réussite" obtenue par la simple division du nombre de candidats ayant réussi l'item par celui des candidats ayant subi l'épreuve: ce sera "R". Une valeur 0.89 est certainement différente de 0.90 et n'importe qui dans les mêmes circonstances attribuerait à "R" la même valeur. "R" est une valeur reproductible.

LA FACILITE EST-ELLE UN CRITERE PEDAGOGIQUE ?

Certainement,et qu'on le veuille ou non.
Soit une épreuve (Sciences Naturelles-1er ou 2éme degré, Anatomie en Mèdecine) composée de deux questions:

1° - Combien y-at-il de doigts à une main ?
2° - Combien y a-t-il d'orteils à un pied ?

La première impression que donne cette épreuve,c'est qu'elle est "facile",n'est-ce-pas ? Celui qui répond 5 et 5 est reçu: si l'épreuve ne comporte que ces deux questions,le candidat n'y est pour rien. Celui qui répondrait autre chose serait refusé; du moins on aime à le croire. Chacun ressent que,dans cette caricature d'épreuve,la moindre faute est impardonnable. Pas tellement,en dernière analyse parce que ces questions sont importantes (après tout,est-il réellement important de savoir combien on possède de doigts...ou de dents ?) mais parce qu'elles sont faciles. Faciles, celà veut dire que l'on s'attend à ce que tout le monde sache répondre. Et si on était déçu ?

Le but de l'apprentissage n'est-il pas de rendre les gestes plus faciles ? Qu'y a-t-il de plus facile : une division ou une multiplication ? Pour des gens qui n'ont jamais appris, les deux opérations sont aussi dificiles (impossibles). Pour les autres, comment répondre ? On soumet un échantillon à une épreuve composée de trente divisions et de trente multiplications comparables quant à leur niveau technique. Si le pourcentage de réussite des multiplications est significativement supérieur à celui des divisions, il sera possible d'affirmer que l'une est plus facile que l'autre. Si on sait,en plus, que la multiplication est apprise avant la division, on pourra tenter d'expliquer cette constatation. Expliquée ou non, la constatation demeurerait. On applique ce test dans une école primaire, à un niveau tel que les élèves soient censés savoir effectuer les deux opérations. Si 80p.100 de la population réussit ses multiplications et 5p.100 ses divisions, on pourra, au moins soupçonner qu'un enseignement a été mieux fait que l'autre. Constater ce fait est une chose, l'expliquer une autre.

LA FACILITE EST-ELLE UN CRITERE DETERMINANT ?

Si le test précédent était un examen,serait-il juste de pénaliser les élèves qui ont eu droit à un mauvais enseignement ?

. NON s'il s'agit d'un contrôle de connaissances, d'un examen courant, c'est-à-dire survenant au cours du curriculum.

. OUI s'il s'agit de recruter des comptables.

Dans le premier cas, la facilité est un critère déterminant de la qualité de l'enseignement : c'est un critère mesurable, déterminant, pédagogique.

LA FACILITE PRISE COMME CRITERE, N'EST-CE PAS UNE SOLUTION DE FACILITE ?

Non, si on veut bien se souvenir que derrière l'apparence de la facilité se cache la réalité de la réussite. Si on attribue à un examen la fonction de distinguer les "insuffisants" et les "suffisants", on peut se poser la question d'une éventuelle altération de l'épreuve si la facilité est prise comme critère déterminant. Deux études ont été faites.

1° - Dans une épreuve comprenant un nombre d'item supérieur à 80 (PHY. KB 6-71) sept ont été choisis : les plus faciles. Le critère de choix était d'avoir été réussi par, au moins, 76p.100 de la population (R = 0.8 arrondi au dixième le plus proche). Choix parfaitement arbitraire mais certainement significatif quant à la facilité puisque moins de 10p.100 des item ont été retenus. Tout aussi arbitrairement il a été exigé au cours d'un examen fictif secondaire, que les candidats aient réussi au moins six sur sept de ces item. L'examen réel avait donné 41p.100 de reçus, l'examen fictif 43p.100. C'est une coïncidence d'autant plus impressionnante que les listes des individus, en gros, correspondaient.

2° - Sur 71 item d'un examen réel (PHY. KB 6-71) il a été recherché leur "pouvoir ségrégatif". C'est-à-dire leur pouvoir de séparation entre les "reçus" et les "refusés". En d'autres termes, il a été recherché leur valeur pronostique. Pour chacun d'eux, un coefficient de ségrégation a été calculé : c'est le rapport des réussites à l'item des "reçus" à l'EXAMEN (à la totalité des épreuves) sur la réussite au même item des non-reçus. Si on représente par "R" les reçus et "NR" les non-reçus, ce rapport a la forme:

$$\frac{R}{NR}$$

. Un rapport = 1 signifie : pouvoir ségrégatif nul puisque reçus et non-reçus réussissent aussi bien.

. Un rapport supérieur à 1 signifie que le pouvoir ségrégatif existe puisque les reçus réussissent mieux.

. Un rapport inférieur à 1 signifie que le pouvoir ségrégatif existe mais qu'il s'exerce à l'envers puisque les non-reçus réussissent mieux.

Le coefficient calculé était retenu comme différent de 1 s'il sortait des limites 1 ± 0.09. Dans l'ensemble, les item les moins ségrégatifs se sont avérés les plus difficiles ($R<0.3$) et ceux d'une extrême facilité ($R>0.9$). On conçoit évidemment qu'un item réussi par tout le monde n'ait aucun pouvoir ségrégatif. Mais il n'était pas évident, à priori, que, les plus difficiles fussent plutôt réussis par les candidats les moins bons.

JUGEMENT BINAIRE ET NUANCES

Chaque item est accepté ou refusé,sans nuances. La base de ce système d'évaluation est rigoureusement binaire. Il est interdit de juger "plutôt bon","excellent","franchement mauvais"ou "je ne sais pas". La correction obéit à la loi du tout ou rien. Le jugement final est-il sans nuance ou,en termes de mesures,dépourvu de sensibilité?.

C'est en 1871 que Henry BOWDICH,fondateur de l'école américaine de physiologie a découvert que le coeur (un grand ensemble de cellules musculaires) ne connaissait que deux états: le repos absolu ou la contraction maximale et,ceci,dans des conditions expérimentales très particulières,sensiblement éloignées des conditions réelles. En 1909 Keith-Lucas débrouille une question très compliquée. Chaque fibre musculaire obéit à la loi du tout ou rien (dans la mesure où on l'excite par une seule stimulation). Hormis l'exception signalée par Bowwdich, un muscle n'obéit pas à cette loi. Or la fibre musculaire est l'unité contractile. Comment donc se fait-il qu'un ensemble (le muscle) n'obéisse pas à la loi régissant chacune de ses unités?. Très simple,répondit Keith-Lucas,à la suite d'une expérience fameuse. Dans un muscle,il y a des dizaines,des centaines,parfois des dizaines de milliers de cellules. Si une seule est excitée,elle répond au maximum et fournit la plus petite réponse possible du muscle: celle d'une unité. Si toutes sont excitées, toutes répondant au maximum et le muscle fournit sa plus grande réponse possible: la somme des réponses de toutes ses unités. Entre les deux extrêmes,autant de réponses peuvent se concevoir qu'il y a de fibres (moins deux). Une contraction musculaire,mesurée en unité de force,est donc la somme des réponses de chaque unité contractile,chacune obéissant à la loi du tout ou rien. La sommation de phénomènes obéissant,chacun, à la loi du tout ou rien,n'est pas un tout ou rien. Il n'est que de comparer la différence de force développée dans l'élévation du bras d'une femme qui se coiffe avec celle de l'haltérophile pour concrétiser la gradation possible de la réponse.

Il en est exactement de même pour l'établissement d'un résultat d'examen: ce n'est pas le résultat lui même qui est un "tout ou rien" mais l'opération unité consistant à accepter ou à refuser chacun des item. La sensibilité dépend alors du nombre d'item proposés et ce nombre est fixé par l'examinateur: le vrai,celui qui rédige l'épreuve.

Dans une contraction musculaire,chaque fibre n'a pas exactement la même valeur d'énergie contractile exprimée en force. Abstraction faite de conditions purement mécaniques (situation,trajet,angle formé avec l'axe osseux...) il n'y a aucune raison pour que toutes les cellules musculaires soient identiques entre elles. Leurs forces se groupent,statistiquement,autour d'une moyenne avec une certaine dispersion. Certaines,ont très peu de force,d'autres beaucoup (relativement). Il est tout à fait possible que deux fibres agissant ensemble ne développent pas autant de force que telle autre agissant seule. Il est encore de même dans l'établissement d'un résultat d'examen: chaque item n'a pas la même valeur et,dans la mesure où la réussite est prise comme critère, rien n'empêche de pousser la division jusqu'à la trentième décimale. (Deux semblent suffisantes).

Au total,cette méthode est sensible,la sensibilité dépendant du nombre d'item et de la précision du coefficient de réussite.

QUALITE DE LA METHODE "R.E.R"

Chacun, dans les mêmes circonstances, attribuera à un item donné, le même coefficient qui n'est rien d'autre que sa fraction de réussite. Un nombre aussi élevé de correcteurs qu'on peut imaginer attribuera la même valeur à n'importe quel item ayant constitué une partie d'une épreuve subie par une population donnée de candidats : cette méthode est fidèle. Sur le document tel que le score d'une épreuve, un examinateur donné trouvera la même valeur pour chaque R élémentaire quelque soit l'intervalle de temps qui sépare l'épreuve de l'établissement de son résultat : il n'y a pas de dérive. Il faut cependant souligner que l'absence de dérive n'est obligatoire qu'en partant du score de l'épreuve et non de l'épreuve elle même. En effet, même pour une épreuve parfaitement constituée mais d'un niveau élevé, il n'est pas certain que tout ce qui était "acceptable" en 1950, l'était encore en 1970. Une nouvelle correction ne pourrait s'imaginer qu'à partir de documents contemporains de l'épreuve. C'est alors l'épreuve qui, éventuellement dériverait mais non la manière d'en juger les résultats.

Sensible et fidèle, cette méthode est-elle juste ? Mathématiquement parlant, parfaitement. Ce qu'on lui demande c'est d'indiquer (+) ou (-) : de dévier du zéro. Instrument sensible et fidèle, son "zéro" sera toujours à la même place : là où on l'a placé et de minimes différences seront révélées. La vraie question est de savoir si la définition du zéro par les candidats eux mêmes est juste.La réponse dépend de ce qu'on attend de l'épreuve. Si on ne sait pas, ce qui s'appelle savoir, comment déterminer l'importance d'un item, si on estime que tous ne doivent pas avoir la même importance, le calcul du coefficient de réussite permet de déterminer une grandeur dans des conditions parfaitement reproductibles. Le coefficient de réussite représente, d'ailleurs, une expression parfaitement fidèle de l'importance: ce qui est bien su, a été bien appris et ce qui a été bien appris l'a été parce que ressenti comme important: ce n'est peut être pas, universellement, le meilleur critère possible de l'importance mais c'en est incontestablement un et, tel qu'il se présente, il n'est pas dépourvu d'avantages.

Il permet d'abord de fort bien cerner le niveau de la population examinée. Pour peu que l'épreuve s'y prête, et il est nécessaire qu'elle s'y prête, un résultat surprenant d'abord peut être compris. Soient, par exemple, deux propositions explorant exactement la même connaissance (telles deux définitions également acceptables du même mot). Si l'une fournit 80p.100 de bonnes réponses et l'autre 20p.100 seulement, on peut immédiatement soupçonner qu'une formulation "passe mieux" que l'autre. Des confrontations successives permettent d'établir les caractères généraux des "mauvaises" formulations. L'enseignant peut alors choisir entre deux attitudes : changer son expression puisqu'elle n'est pas comprise ou faire en sorte qu'elle devienne compréhensible. Ainsi l'épreuve aura révélé, indirectement, mais indiscutablement, un défaut dans l'enseignement d'une part (car c'est un défaut d'être mal compris), et d'autre part l'influence de la forme d'une question sur la grandeur du résultat.

Celà peut encore aller beaucoup plus loin. En imaginant que deux professeurs soient en désaccord total, dans le même établissement

sur la méthode à employer pour enseigner les langues, chacun ayant de solides arguments, il est possible de s'en remettre à l'examen pour trancher. Si l'un fait apprendre à ses élèves un lexique et la grammaire et si l'autre utilise, grâce au film, des "tranches de vie", une épreuve peut être conçue en deux parties. L'une sera une épreuve de traduction de mots, l'autre une épreuve de conversation.

Si les candidats du professeur-dictionnaire ont de meilleurs résultats en traduction et ceux du professeur audio-visuel de meilleurs résultats en conversation, il sera possible d'établir les mérites respectifs des deux méthodes (puis de choisir, après qu'aura été décidé "pourquoi on apprend des langues). Mais des deux groupes de candidats, aucun ne sera favorisé. Ceux du professeur-dictionnaire ayant, par hypothèse, mieux réussi les épreuves "de mots",que ceux du professeur audio-visuel, dans le premier groupe un échec coûtera plus cher et un succès rapportera moins que dans le second. Cette méthode est ouverte à toutes les expériences. Elle permet de mesurer l'efficacité d'une méthode sans que n'en pâtissent les candidats.

Pour sauter de la justesse à la justice, la question à laquelle il est bien difficile de répondre concerne la définition du "modèle". Est-il juste que la "vox populi" le détermine ? Il semble bien que, dans nombre de cas, l'acceptation de cette solution constitue un progrès dans la mesure, au moins, où celà permet de définir un modèle sur des critères précis. Avant de refuser énergiquement d'adopter le coefficient de réussite comme juge unique et arbitre, il faut se demander si ce qu'il remplacerait était réellement meilleur.

Enfin, il faut souligner, avec force, que connaître la population enseignée et son degré d'évolution permet une réelle correction. C'est dire que cette méthode permet et justifie un contrôle continu où les résultats, considérés comme ce qu'ils sont: des informations, permettent ce qu'on appelle souvent un "feed-back" et qui, en réalité,est une régulation.

Au total, cette méthode est sensible et fidèle. Parfaitement juste quand il s'agit d'évaluer une progression ou de tester des méthodes sur leurs résultats, elle ne peut suffir, à elle seule, à définir la qualité du produit fini. Une balance peut peser n'importe quoi, on ne peut pas en attendre qu'elle renseigne sur la nature du produit pesé. Il semblerait plus juste de choisir la valeur de chaque item: que la méthode d'établissement des résultats ne soit pas fondée sur la réussite mais sur un critère extérieur: le coefficient NORMATIF.

APPLICATION DE COEFFICIENTS NORMATIFS

Ce sont des coefficients imposés à priori. Tel item sera affecté d'un coefficient (1),tel autre, d'un coefficient (0.5)...etc. Quels sont, dans l'application stricte de la méthode proposée, les conséquences de leur utilisation ?

Deux cas schématiques sont à considérer: un cas simple où tous les item ont le même coefficient, un cas complexe où les coefficients n'ont pas tous la même valeur.

TOUS LES ITEM SONT AFFECTES D'UN MEME COEFFICIENT.

Celà signifie que,pour l'examinateur, tous les item ont la même importance et que cette importance même est exprimée par un nombre significatif. La rigidité de cette introduction peut être sensiblement atténuée par une autre formulation. Un examinateur juge, éventuellement après consultation, qu'il est réellement impossible de faire la moindre différence d'importance (jugée, par exemple, sur des critères d'utilité) entre les propositions qu'il soumet à ses candidats. La valeur proposée du coefficient exprimera, en fait, la proportion d'erreurs admissible. Si chaque item est coté (1), celà signifie que, quelqu'en soit le nombre, une erreur est rédhibitoire. Affecter ses item d'un coefficient (0.9) signifie pour l'examinateur, qu'il propose ce contrat: "il vous est reconnu le droit de mal répondre une fois sur dix". Plus directement, le coefficient exprime le degré d'exigence: exigence de 100p.100 pour N = 1; de 90p.100 pour N = 0.9,....de 10p.100 pour N = 0.1. Ainsi, N = 0.98 signifie que la marge tolérée est de deux pour cent: l'épreuve doit être composée, au moins, de 49 item, une seule erreur étant admise, ou d'un nombre d'item multiple de 49, le nombre d'erreur admises croissant proportionnellement. Il est évidemment inadmissible de composer une épreuve de 60 item avec un N de 0.98: accepter une seule erreur donnerait à N une valeur de 1/60 (0.9834); en accepter deux, une valeur de 1/30 (0.9667) ,toutes deux différentes de la valeur déclarée.

LES COEFFICIENTS N'ONT PAS TOUS LA MEME VALEUR.

Soient deux catégories d'item: les uns cotés (1), les autres (0.1). Si on suppose une épreuve composée, notamment,de 20 item (1), un candidat qui en aurait réussi 19 aurait le bilan provisoire suivant

. Actif: $N\,(1-N) \times 19 = 1(1-1) \times 19 = 0$

. Passif: $-N^2 \times 1 = -(1)^2 \times 1 = -1$

Bilan $= -1$

Chaque item (0.1) réussi inscrit à l'actif: 0.1 (1-0.1) = + 0.09.
Pour"compenser" l'échec dans un item (1), il faut réussir un nombre de (0.1) de $\frac{1}{0.09}$ soit: 12 (11.111 est plus grand que 11).
Si on admet une telle compensation il faut donc proposer 12 item (0.1); le bilan d'un candidat ayant réussi 19/20 des item (1) et 12/12 des item (0.1) serait de - 1 + 0.09 x 12 = - 1 + 1.08 = + 0.08: il serait reçu.

Il n'est pas certain, cependant, qu'une telle réussite soit logique. En effet:

. Il a manqué un item (1) pour lequel on avait ainsi affirmé l'échec intolérable.
. Il a réussi 12 item (0.1) d'importance presque négligeable: il était proclamé suffisant d'en réussir 1 sur 10 soit 2 sur 12. L'opération revient à considérer que les item très importants (1) ne le sont pas autant qu'on le dit: on tolère, en fait, un échec sur vingt ce qui peut s'exprimer par un coefficient normatif N = 0.95. On peut de même abaisser le coefficient proclamé (1) à 0.90 en introduisant dans l'épreuve non pas 12 mais 24 item (0.1) ... etc.

Au total, le coefficient "N" peut parfaitement être appliqué au calcul des résultats. Son utilisation exige une précision d'autant plus complexe que l'épreuve "normative" est moins homogène.

LE MANIEMENT DE CES COEFFICIENTS N'EST PAS SIMPLE.

Soit un item normatif 0.8. Sa définition appelle deux questio

1ère question: Que vaut 0.8 par rapport à 0.4 ?

On pourrait penser que 0.8 représente une valeur double de 0.4. En terme d'épreuve, celà peut se traduire ainsi:

. Un échec en (0.8) équivaut à deux échecs en (0.4): c'est vrai.
. Un échec en (0.8) est compensé par deux réussites en (0.4): c'est faux.
. Une réussite en (0.8) compense deux échecs en (0.4): c'est faux.

Soit une épreuve composée de trois item: a (0.8), b (0.4), c (0.4). Deux candidats A et B ont respectivement:

A: réussi b et c, manqué a.
B: réussi a , manqué b et c.

Résultat de A: $-(0.8)^2 + 0.4\,(1-0.4) + 0.4\,(1-0.4)$
$- 0.64 + 0.4 \times 0.6 + 0.4 \times 0.6$
$- 0.64 + 0.24 + 0.24$
$- 0.64 + 0.48 = -0.16$ Refusé

Résultat de B: $+ 0.8\,(1-0.8) - (0.4)^2 - (0.4)^2$
$+ 0.8 \times 0.2 - 0.16 \times 2$
$+ 0.16 - 0.32 = -0.16$ Refusé

Donc un échec (0.8) vaut deux échecs (0.4)
Mais:

. Un échec (0.8) n'est pas compensé par deux réussites (0.4): A
. Une réussite (0.8) ne compense pas deux échecs (0.4): B

Pour permettre au candidat A (échec en 0.8) d'être reçu, il eût fallu que l'épreuve fût composée autrement:

. Quatre autres item (0.8) ou
. Trois autres item (0.6), (0.5), ou (0.4) ou
. Quatre autres item (0.3) ou (0.2) ou
. Huit autres item (0.1) ou (0.9).

Au total, coter un item (0.8) ne peut pas signifier qu'il en vaut deux cotés, chacun, (0.4).

2ème question: Que vaut 0.8 par rapport à 0.7 ou à 0.9 ?

Il faut simplement être capable de répondre,"il est sûr que le (0.8) a plus d'importance que le (0.7) et moins que le (0.9)" sans craindre d'être démenti par les faits ou par l'opinion unanime de collègues aussi compétents.

"AVANTAGES" D'UN CRITERE NORMATIF DE REUSSITE

C'est typiquement le genre de sujets dans lequel il est nécessaire de discerner les apparences des réalités. En principe, le choix de ce genre de coefficient se justifie parfaitement dans la mesure où l'examen est censé fournir une sanction et non une correction, par exemple, dans un examen terminal ou dans un examen de qualification professionnelle. Il est tout à fait possible,dans ces conditions, de composer une épreuve uniquement avec des item importants. Mais, l'esprit est prompt.... Qu'adviendra-t-il réellement des belles résolutions à la minute de vérité: lorsqu'on constatera qu'un très faible pourcentage des candidats a satisfait aux exigences ? Ne fera-t-on, en jury, le procès de l'examen ? Ce ne serait pas une mauvaise chose mais il est à craindre que l'opération ne s'effectue pas dans les conditions requises de sérénité. Il est à peine croyable qu'un jury décide de se donner huit jours de réflexion (il s'agirait bien de huit jours de réflexion et non du renvoi à huit jours d'une décision désagréable!). Il est plus probable qu'une improvisation hâtive "arrange" les résultats avec le double inconvénient d'une décision fortement aléatoire et d'une infidélité à une ligne de conduite. En vérité, il faut être sûr de soi pour oser fixer un coefficient "à priori" et il n'est pas exclu qu'une telle certitude soit synonyme de témérité.

Dans un examen de Mèdecine, à un certain niveau, exiger une réponse juste à des questions importantes ne semble pas du tout trahir le sens français de "contrôle". On peut prendre comme critère, par exemple que l'erreur qui tuerait le malade doit tuer le candidat. Il est tout de même quelques domaines où la différence est nette entre ce qui doit et ne doit pas se faire. Mais un jury peut "fondre" in extremis. Est-on vraiment sûr que ce que dit le candidat tuerait le malade ? On peut encore rater son homme en lui tirant un coup de revolver dans la tête! Et il est vraisemblable que la tentative de meurtre symboliquement commise par le candidat ne serait pas aussi franche. Le jury redeviendrait la proie de ses démons et plus particulièrement, de celui du scepticisme. Pour cette raison, à laquelle s'ajoute éventuellement la considération qu'un lapsus peut arriver, choisir un coefficient "1" ne semble pas réaliste. Il est bien difficile de donner une valeur "pas trop éloignée de 1" qui, arbitraire par définition, ne soit pas parfaitement subjective.

Il serait hautement souhaitable que certains examens soient normatifs et il est tout à fait possible qu'ils le soient. Leur réalisation exige une certaine force et des connaissances qui ne semblent pas très répandues. Tant que la plupart des examinateurs se déclarent incapables de justifier la constitution de leur épreuve, aussi longtemps qu'on ne saura pas prévoir le caractère facile d'une question, jusqu'à ce que toutes, absolument toutes les réponses attendues d'une épreuve soient parfaitement prévues, en bref tant qu'une épreuve n'est pas rigoureusement définie par son but, il n'est pas honnête de présenter comme normatif ce qui ne serait que despotique. Lorsqu'une épreuve est normative il faut qu'elle soit annoncée comme telle, que ses critères en soient définis et publiés et, son texte ayant été préalablement contrôlé, que la décision prise soit appliquée aussi bien si elle fournit un effectif de reçus de 1 que de 99p.100. Si tous les "normatifs" savaient que, l'épreuve étant conçue jusque dans ses coefficients, il était absolument impossible d'en modifier quoique ce soit, peut être y en aurait-il moins ?

Certifier qu'un candidat reçu possède toutes les qualités qu'on attend de lui, suppose qu'une vérification qualitative et quantitative en a été faite. Il n'est alors de bon examen que normatif. Si on ne sait pas le faire, on peut apprendre. Une assez bonne méthode consisterait à passer un premier stade: pragmatique. Beaucoup de défauts, tant de l'enseignement que de l'examen, pourraient être mis en évidence. L'imbécilité consistant, non à faire des erreurs, mais à y persister lorsqu'on les a reconnues, ce défaut étant rare chez les enseignants (peut être même absent) d'examens pragmatiques, il résulterait nécessairement une amélioration de l'enseignement et de ses moyens de contrôle et on pourrait alors devenir normatif si celà restait nécessaire toutefois, car un excellent enseignement et de bons examens fourniraient peut être des coefficients de réussite proche des coefficients normatifs. Ce qui ne voudrait pas dire que tout le monde serait nécessairement reçu: toutes les questions réussies à 90p.100 ne signifient pas que chacun les a réussies à 90p.100. Une collection, patiemment constituée d'item réussis à 90p.100 permettrait de savoir qu'il est, en toute certitude, admissible de les exiger puisqu'il aurait été démontré que la performance exigée était réalisable, utile et démonstrative.

FAUT-IL ETRE NORMATIF OU PRAGMATIQUE ?

Tout dépend de ce qu'on attend d'un examen. S'il s'agit d'un contrôle terminal dont la sanction est une qualification professionnelle, il vaudrait mieux être normatif. Il est franchement souhaitable de l'être. Encore faut-il pouvoir, et c'est relativement facile par le jeu du coefficient, et savoir. Incontestablement, un mèdecin doit avoir reçu une estampille de qualité fiable qui lui permette réellement de prendre les lourdes responsabilités qui seront les siennes. Ce dont le mèdecin n'est pas responsable c'est de la fiabilité de l'estampille.Les enseignants, oui. Deux questions se posent donc: en quoi et comment se montrer normatif si on a la volonté de l'être ?

COMMENT ? C'est assez facile. Très facile en ce qui concerne la mesure dans une épreuve homogène: on impose un coefficient élevé aux item. Il est même possible d'imposer le coefficient maximum: 1 à chacun. Ainsi,tout item manqué vaudra - 1 et tout item réussi,zéro. Une seule erreur est éliminatoire. On peut légèrement adoucir si on estime devoir tenir compte de lapsus toujours possibles: un "N" de 0.95 ne permet pas tout à fait la compensation d'un échec par 21 réussites;l'épreuve doit alors comporter au moins 23 item sans quoi il ne sert à rien d'abaisser l'exigence de 1 à 0.95. C'est un peu moins facile en ce qui concerne les épreuves non homogènes.

EN QUOI ? C'est beaucoup plus difficile tant qu'une liste "d'objectifs normatifs" ne sera pas établie. Celà demanderait du temps, de la réflexion, de très nombreuses informations. Cette liste n'existe pas (n'a jamais été communiquée) et mériterait d'exister. Elle serait alors normative des normatifs.

La position de pragmatique, critiquable dans une épreuve de qualification, est largement défendable dans toutes les autres circonstances. Le pragmatique étant l'examinateur qui attend le score pour établi ses résultats: son coefficient sera "R" ou coefficient de réussite. Ce qui implique qu'il accorde plus d'importance à ce qui a généralement été réussi. On pourrait dire: que le pragmatique est réaliste.

Réaliste, parce qu'il avoue faire ce que chacun fait sans l'avouer: noter en fonction du "candidat moyen". Cet aveu fait, coûteux peut-être, il commence par réellement connaître le "candidat moyen". Sur des chiffres, pas sur des impressions. C'est, au moins, plus précis. C'est aussi plus clair: ce critère sera réellement le seul.

Réaliste, parce qu'il avoue ce que peu avouent: dans la détermination de l'importance des item, il rentre une part considérable, variable et inexprimable par des chiffres,de facteurs personnels. Personnels aux correcteurs, variables d'un correcteur à l'autre et, pour un même correcteur, d'un moment à l'autre. Ce n'est pas un critère fiable pour un moyen de contrôle.

Réaliste, enfin, parce qu'il considère que, dans de mauvais résultats, l'enseignement et la forme même de l'examen peuvent être pour quelque chose et qu'il en tient compte. Il ne dit pas: "évidemment, l'enseignement n'est pas parfait, évidemment, l'examen n'est pas parfait mais les candidats, non plus, ne sont pas parfaits". Il ne dit pas ce qui ne

veut rien dire. Une faute d'enseignement (non sa nature), une faute d'examen sont dénoncées. Défalcation faite de ces fautes, le reste appartient aux candidats. Certainement.

On est souvent pragmatique ou normatif par nature. Il faut savoir être l'un ou l'autre par sagesse, même en allant "contre sa nature". Celà n'est-il pas de l'objectivité ?.

COUTUME ET REALISME

Reprenons notre cariture d'épreuve: combien y a-t-il de doigts à une main et d'orteils à un pied ? La question est de savoir quel score sera celui d'un candidat qui aurait répondu 5 et 4. Un certain nombre d'étudiants à qui ce problème a été soumis a répondu étourdiment (étourdiment ?) 9 sur 10. Donc un point par article ? Ceci mérite réflexio car cette réponse a été la plus fréquente. Peut-être peut-on en déduire qu'il s'agit d'une décomposition spontanée en item. Mais tous les item ont la même valeur. Celà peut encore révéler qu'un item est analytique: doigt par doigt, et non synthétique: une main ou un pied. Avec quelques uns (échantillon de représentativité non garantie) le score a été discuté: "si vous ne savez pas combien une main comporte de doigts, il vous manque quelque chose d'important dans la connaissance que vous prétendez en avoir "Certainement, répondirent-ils, mais il aurait fallu que les questions fussent posées "autrement" pour que la sanction que vous proposez soit juste". Cet "autrement" n'a malheureusement jamais pu être défini (nous ne sommes pas examinateurs). Il semble donc que refuser notre candidat "5 et 4" n'est pas une sanction d'une évidente justice. Le coutumier serai bien embarrassé et il mettrait vraisemblement 5 sur 10. S'il a des remords devant cette note exagérée, il est réaliste.

Le réaliste normatif refuserait le candidat sans hésiter et on éprouverait quelque peine à lui démontrer qu'il a tort. Et pourtant si, par impossible, l'information n'avait jamais été communiquée aux étudiants, aujourd'hui candidats ?

Le réaliste pragmatique chercherait la réponse. Combien de candidats ont répondu autre chose que 5 et 5 ? Un pour cent ? L'information avait été donnée. Quatre vingt pour cent ? Elle n'avait pas été donné En fait, on peut supposer que la plupart des individus connaissent la réponse, pour peu qu'ils sachent compter. La population aura répondu dans sa grande majorité "5 et 5". Le malheureux 1p.100 qui aurait répondu "5 et 4" aurait un S.P.I. précis.

1°- Tout le monde a répondu 5 à la première question. R = 1. Chacun a zéro (c'est un score suffisant).

2°- 99p.100 ont répondu 5 à la seconde: R = 0.99. Tous ces candidats ont donc: 0.99 (1-0.99) = + 0.01

3°- Le malheureux 1p.100 qui a répondu autre chose (réponse refusée qu'il s'agisse de 4,3,2,1 ou 0) a,pour cet item,$-(0.99)^2$ soit -0.98

4°- Au total:

99p.100 des candidats ont 0 + 0.01 = + 0.01: Reçus

1p.100 des candidats a 0 - 0.98 = - 0.98: Refusés

Pragmatique et normatif fournissent la même sanction: réalist Le coutumier risque de conserver ses remords et de recevoir son candidat, faute d'un critère défini lui permettant de faire autrement.Encore n'aura t-il de remords que s'il est réaliste! Le pragmatique,seul, sait exactement pourquoi et comment il établit son résultat. Ne semble-t-il pas?.

CONCLUSION

La méthode de détermination des résultats, dont le mérite de la formulation revient à G.BARRES, présente un certains nombre d'avantages qui, semble-t-il, compensent largement l'inconvénient réel que représente le temps nécessaire aux calculs. Mais... on n'a rien sans rien.

L'utilisation du coefficient de réussite représente une méthode objective, établie sur une réalité, facilement reproductible, sensible,fidèle et sans dérive permettant toutes les audaces dans les méthodes d'enseignement et d'examen sans que les candidats n'en souffrent. La nécessité d'établir un score fournit un matériel d'une considérable utilité en vue de l'analyse de l'examen lui même aux fins de déterminer, notamment, sa valeur en tant que moyen de contrôle donc sa propre critique L'utilisation d'un coefficient normatif laisse un peu plus réticent non que la méthode ne soit pas applicable mais certains doutes dans l'actuel état des choses, peuvent être opposés à la validité d'une définition acceptable d'un tel coefficient.

Précisément, l'utilisation du coefficient de réussite et les informations critiques qu'elle peut fournir représente une étape utile dans l'approche d'une forme d'application qui, dans certains cas au moins, semble plus juste.

Au total, par la précision et l'objectivité qu'elle apporte dans un domaine qui n'en est pas richement pourvu, cette méthode est nécessaire. Elle n'est pas suffisante car des résultats mathématiquement parfaits n'ont de sens que si l'épreuve elle même en a.

CHAPITRE III

DU RESULTAT A LA NOTE

DU RESULTAT A LA NOTE 216
PREMIER TEMPS: ETABLISSEMENT DU SCORE 217
DEUXIEME TEMPS: PONDERATION DU SCORE 219
TROISIEME TEMPS: NOTATION 221
RESUME 222
QUELQUES EXAMENS FICTIFS 223
1er Exemple 224
2e Exemple 225
3e Exemple 226
4e Exemple 227
5e Exemple 228
Résumé 229
DES GRANDEURS INTERESSANTES : DIFFERENCE DES VALEURS OBTENUES SELON LA METHODE UTILISEE 230
Amplitude 231
Positif maximum attribuable 233
Négatif maximum attribuable 234
SIGNIFICATION DES DIFFERENCES 235
DES VARIANCES OBTENUES PAR SOUSTRACTION 238
QUE REPRESENTENT CES VARIANCES ? 239
SIGNIFICATION DES GRANDEURS A, P, N, $\overline{R}$ 240
ET LES CANDIDATS ? 242

Dans la mesure où l'attribution d'une note est utile il faut que la valeur qu'elle exprime soit significative et objective. La seule façon de rendre une note significative est de définir ce que veulent dire les différents degrés de l'échelle. La seule façon de la rendre objective est de la définir par rapport à l'objet étudié, ici une population de candidats. La méthode présentée consiste, à déterminer la note à partir des résultats: elle procède à l'inverse des habitudes.

Les résultats étant établis à partir du score, l'analyse soigneuse du score permettra de découvrir qu'il est finalement très simple de donner à une note une signification très claire: en dehors de la moyenne, ici rigoureusement définie, toute autre valeur représente un éloignement par rapport à la référence dans des limites parfaitement définies qui sont celles de la distribution des résultats autour de la moyenne (variance).

Après l'examen attentif du score et quelques questions fondamentales sur la notation, l'étude de cinq examens fictifs, élaborés pour leur valeur démonstrative, montrera qu'il n'est pas nécessaire d'être grand mathématicien pour fonder la notation sur l'unité de dispersion: l'écart-type, puisque le score a la bonne grâce de fournir des variances par simples soustractions. En possession de grandeurs significatives et dont la signification sera démontrée, il sera possible de passer au chapître concret "une notation normalisée".

La note exprimant finalement un degré d'éloignement par rapport à la moyenne, la moyenne représentant le candidat ayant réussi "à minima", c'est le coefficient de réussite qui sera pris en considératio

DU RESULTAT A LA NOTE

Si on estime utile de "mettre une note" c'est qu'on veut traduire en un langage unique des résultats primitivement différents en résultats formellement comparables. Pour celà on utilise une échelle, formée d'un certain nombre de degrés. Par tradition, et souvent par obligation réglementaire, l'échelle utilisée est vigintésimale, ce qui signifie comme ne le dit pas l'étymologie, qu'elle a vingt et un degrés. Si on exprime des résultats en degrés, il est absolument indispensable que la situation sur un degré plutôt que sur un autre ait une signification. Une signification de valeur, de préférence: que la situation sur un degré supérieur traduise une valeur supérieure et que cette distinction même soit significative. C'est dire qu'il faut définir ce que veulent dire: vingt ou onze, ou dix, ou zéro, ou n'importe quel échelon. En vérité, l'échelle des valeurs d'un examen devrait être plus comparable à celle des musiciens qu'à celle des températures: l'intervalle séparant deux degrés n'est pas nécessairement le même partout. Ainsi, sensible et tonique sont deux degrés parfaitement définis si on connait le mode; sus-dominante est un autre degré aussi parfaitement défini. Cependant, en majeur, l'intervalle qui sépare sensible de tonique est plus petit que celui qui sépare la sus-dominante de la sensible. Ces trois degrés étant pourtant consécutifs.

L'objectif de la notation étant de traduire aussi rigoureusement que possible un score en note, il faut utiliser les données significatives du score.

PREMIER TEMPS : ETABLISSEMENT DU SCORE

Il est simple, quelques conditions impératives étant respectées.

1°- Un catalogue complet des item doit être dressé avant de regarder la première copie. Les centres d'intérêt qui seront pris en considération dans la correction doivent être rigoureusement inventoriés. Ceci avant tout et sans aucun souci au sujet du pronostic. On ne doit pas tenir compte du fait qu'on pense qu'il serait possible que beaucoup de candidats ne répondent pas sur tel ou tel point. Une seule préoccupation: la réponse recherchée correspond-elle à la question ? Répondre oui, implique que le point en question soit un item.

2°- Le nombre d'item est sans importance pour peu qu'il soit assez grand (au moins trente). En particulier, il est parfaitement vain de chercher à tout prix un nombre "juste": par exemple 40 si on note sur 20. "Juste" ou "pas juste" ne change rien aux calculs qu'il faudra bien faire. Ce nombre, fixé à priori et sur des motifs purement pédagogiques ne devra plus être modifié, quelles que soient les impressions suscitées par la lecture.

3°- Etablir le score se bornera exclusivement à "pointer" ce qui est accepté. Le pointage, avec un autre signe, de ce qui est refusé peut servir de contrôle, chaque "copie" devant naturellement comprendre le même nombre d'item.

4°- Dans la mesure où la forme de l'épreuve s'y prête, il y a avantage à cocher non pas les copies une par une, mais un certain item, ou un groupe d'item donnés, à la suite. C'est l'item qui est l'unité, pas la copie. Si l'épreuve est constituée de deux "questions" rédigées, il est préférable de corriger les questions une par une: toutes les premières, puis toutes les secondes (ou inversement).

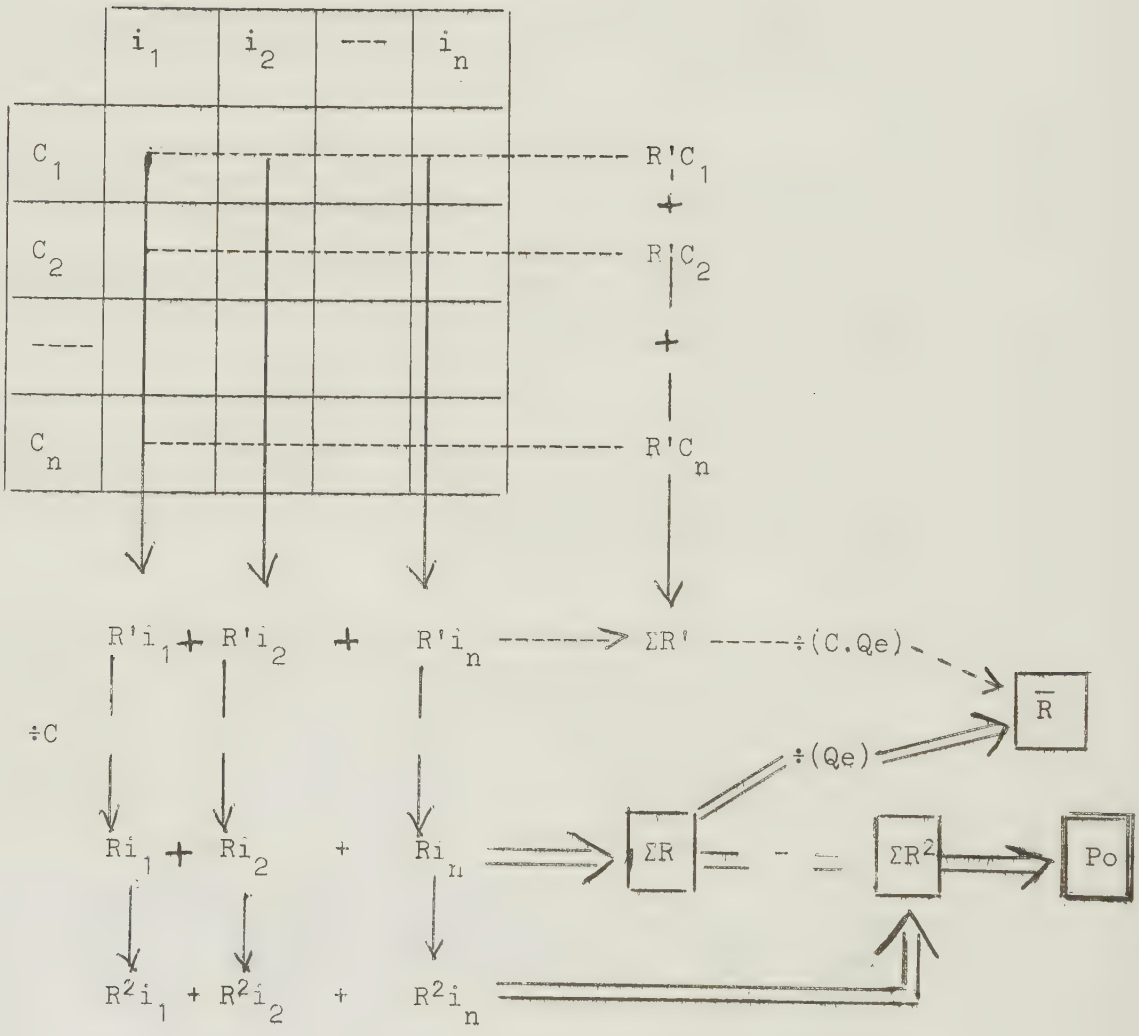
Au total, établir le score n'implique qu'un jugement binaire sur chaque item: "accepté" ou "refusé", (+) ou (-), (1) ou (0). C'est rigoureusement tout: les réponses non faites étant évidemment refusées.

UN SCORE :

Item N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
C 1	I	I	I	I	I	I	I	I	I	I	I		I	I	I	I				I	I	I	I	I						
C 2	I	I	I	I	I	I			I	I	I		I	I	I		I					I	I	I		I	I			
C 3	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			I	I	I		I	I		I	I	I	I	I
C 4	I	I	I	I	I	I			I	I	I	I	I	I		I	I		I	I	I	I				I	I			
C 5	I		I		I	I		I	I	I	I	I		I	I				I	I	I		I	I		I	I			
C 6		I					I	I	I	I	I	I	I	I	I		I		I	I	I		I	I			I	I		
C 7	I						I	I	I	I	I	I		I	I						I	I	I							

UTILISATIONS DU SCORE

Pour C candidats et Qe item



RESULTAT DU CANDIDAT n $\quad \Sigma R_{cn} - \Sigma R^2$ $\quad \geq 0$ RECU (+) $\quad < 0$ REFUSE (-)

VARIANCE DES CANDIDATS : Po $\quad \sqrt{Po}$ = Unité de NOTATION

VARIANCE DES ITEM : $\dfrac{\Sigma R^2 - Qe.\bar{R}^2}{Qe - 1}$: base de l'analyse d'enseignement

DEUXIEME TEMPS : PONDERATION DU SCORE

C'est le temps le plus complexe.

1°- On dresse le tableau complet des scores en binaire (1 ou 0). Ce tableau est composé de colonnes (verticales) et de rangées (horizontales):

a) chaque rangée représente un candidat.

b) chaque colonne représente un item.

2°- On compte dans chaque colonne le nombre d'acceptés (nombre de + par exemple): ce qui fournit le nombre (R' de réponses justes fournies à l'item (i).

3°- On divise le nombre R'i par C, on obtient ainsi R_i: le coefficient de réussite de l'item considéré. R est représenté par une valeur comprise entre 0 et 1. En fonction de la sensibilité recherchée on retiendra un nombre de décimales plus ou moins élevé. Si on souhaite classer rigoureusement les candidats les uns par rapport aux autres, deux décimales suffisent pour un effectif de 300 avec 30 item. En effet, 30 item celà fait 30 échelons possibles avec un seul chiffre significatif (la première décimale). Avec deux chiffres significatifs (deux décimales) celà représente 300 échelons. En règle, deux décimales suffisent.

4°- (FACULTATIF). On compte le nombre de (+) rangée par rangée: on obtient ainsi le score individuel exprimé en nombre d'item (résultat non significatif).

5°- (FACULTATIF). On fait la somme des réussites individuelles on trouve ainsi le score total (nombre d'item réussis au total: par tous les candidats).

6°- (FACULTATIF). On fait la somme des totaux de chaque colonne (R'), on trouve ainsi le score total de l'ensemble des candidats qui doit évidemment être le même que précédemment (vérification).

7°- (FACULTATIF). On divise le score total par le produit du nombre de candidats et du nombre d'item que comportait l'épreuve on obtient ainsi la réussite moyenne $\overline{R}$.

$$\overline{R} = \frac{\Sigma R'}{C.Qe}$$

8°- On porte chaque valeur de R au carré: le temps nécessaire est à peu près le même, qu'on utilise une table de carrés ou une machine à calculer. Il n'est indispensable de retenir que deux décimales. Ainsi sont connus tous les R^2 qui sont les négatifs.

9°- On calcule (1 - R)R pour chaque item: ce sont les positifs.

10°-On fait la somme, pour chaque candidat, de ses positifs et de ses négatifs: ainsi est établi le résultat individuel significatif. Le résultat proprement dit est fourni par le signe de la somme: (+) pour RECU et (-) pour RECUSE. La valeur de la somme représente le Score Pondéré Individuel, base de la NOTATION.

11°- ON PEUT SIMPLIFIER

Le positif s'exprime par $R(1 - R)$ et le négatif par R^2. Le positif peut encore s'écrire sous la forme: $R - R^2$. Ainsi le positif apparaît comme étant la grandeur R calculée en (3°) diminuée du négatif R^2 (calculé en 8°). Le résultat d'un étudiant peut se représenter par $\Sigma R - \Sigma R^2$. Que l'item soit réussi ou non, on utilise la grandeur R^2 (négatif). Si l'item est réussi, et dans ce cas seulement, on ajoute à R^2 la valeur de R. Calculer le positif est donc inutile.

a) Une fois pour toutes on calcule ΣR^2 qui représente le négatif maximum et on l'attribue à chaque candidat: c'est, en quelque sorte, le handicap,(représenté plus loin par N pour négatif).

b) Pour chaque candidat on fait la somme des R, chacun pouvant avoir une valeur différente selon l'item considéré. Chaque candidat sera donc caractérisé par une somme de R, fidèle reflet des item qu'il a réussis.

c) Pour chaque candidat on soustrait de la somme des R, (différente d'un candidat à l'autre) la somme des R^2 (ΣR^2, valeur constante pour l'épreuve).

Finalement, trois résultats sont possibles:

1°- $\Sigma R - \Sigma R^2 > 0$: candidat reçu.
2°- $\Sigma R - \Sigma R^2 < 0$: Candidat refusé.
3°- $\Sigma R - \Sigma R^2 = 0$: candidat douteux (en fait :reçu ce qui reste à établir)

TROISIEME TEMPS : NOTATION

Toute une partie de ce chapitre en traitera en détail. Pour le moment, quelques faits méritent examen à propos de la notation.

1°- Le résultat est-il une notation ?

NON, parce que sa valeur n'est que relative, elle dépend de l'épreuve et son écart n'est pas prévisible.

2°- La notation est-elle utile ?

NON, si la sanction de l'épreuve s'exprime en binaire: refus ou reçu. Le seul signe du résultat suffit à répondre à la question (zéro étant considéré comme le plus petit positif).

NON, si le but de l'épreuve est de classer les candidats (composition traditionnelle ou concours).

OUI, si le but de l'épreuve est de permettre des comparaisons:

- avec des épreuves de même nature se suivant dans le temps
- avec des épreuves simultanées de natures différentes composant un même examen.

OUI, si le but de l'épreuve est d'aboutir à un résultat chiffré destiné à entrer dans une moyenne entre épreuves de même nature ou de natures différentes.

3°- A quoi sert la notation ?

A exprimer un résultat dans une forme préalablement définie (échelle vigintésimale habituelle, décimale ou centésimale, parfois octogésimale si l'expression sur 20 se fait au quart de point). Elle sert de langage commun, conventionnel,à différentes épreuves: que celles-ci soient ou non de même nature et quel que soit l'intervalle de temps qui les sépare.

Passer du résultat à la note, c'est faire une traduction qu'il n'y a lieu de faire que si elle est utile.

RESUME

I - Etablissement du score, par tout ou rien d'aprés un modèle pré-établi. Il revient dans tous les cas à l'examinateur d'établir le modèle. L'établissement du score peut consister à faire le tableau si la correction est manuelle Si la correction est automatique la confection du tableau est remplacée par la programmation de la machine. La correction est obligatoirement manuelle si les questions (brèves ou longues) sont ouvertes, car la machine ne peut pas lire les synonymes ni comprendre les mots mal orthographiés et les réponses devraient être dactylographiées. Toute l'opération ne peut être automatique que dans une épreuve uniquement constituée de questions fermées.

II - Pondération du score:
1°- tableau (qui devient le seul instrument. Une machine peut le traîter, toute la suite peut être automatique)
2°- calcul du R par item (sommes des réussites divisée par nombre de candidats).
3°- élévation au carré de chaque R.
4°- addition des R^2 (une fois pour toutes).
5°- somme des R pour chaque candidat.
6°- différence pour chaque candidat: somme des R - somme des R^2. Résultat établi par le signe de la différence. La grandeur de la différence constitue le Score Pondéré Individuel (S.P.I).

III - Du Score Pondéré Individuel à la note.

QUELQUES EXAMENS FICTIFS

L'avantage de la fiction sur la réalité, c'est qu'on peut imaginer des résultats que l'on souhaite afin,de faire apparaître des grandeurs significatives dont la variation volontaire permet de saisir la signification. Dans la réalité, il serait bon de parvenir à comprendre ce qui s'est passé à partir des données qui ne seront utiles que si elles sont comprises. Le but de ce qui suit est de justifier le caractère nécessaire d'un certain nombre de calculs et de montrer qu'il ne s'agit pas de broyer des chiffres pour le seul plaisir d'utiliser un jargon.

ABREVIATIONS UTILISEES.

Qe : Nombre d'item de l'épreuve.

C : Nombre de candidats ayant participé à l'épreuve.

$Qe \times C$: Nombre d'item jugés.

R' : Nombre de réussites.

$R'_{.in}$: Nombre de candidats ayant réussi le nième item.

R'_{cn} : Nombre d'item réussis par le nième candidat.

$\Sigma R'$: Nombre total d'item réussis par l'ensemble des candidat

R : Coefficient de REUSSITE d'un item :

$$R = \frac{R'}{C}$$

$\overline{R}$: Réussite moyenne: $\frac{\Sigma R}{Qe \times C}$

ΣR : Somme des coefficients de réussite de chacun des Qe item de l'épreuve.

R_{i1} : Coefficient de REUSSITE de l'item N°1.

PREMIER EXEMPLE. (Il sert notamment à l'introduction des abréviations.)

Données :

. Nombre d'item de l'épreuve : ..... Qe = 40

. Nombre de candidats ayant subi l'épreuve :....C = 100

. Nombre de réponses acceptées $\Sigma R'$..... 2000

. Nombre d'item jugés :..... Qe x C = 4000

. Chaque item a été réussi par 50 p.100 des candidats :

$R_{i1} = R_{i2} = R_{in} = 0.5$

Déductions :

1°- La réussite moyenne ($\bar{R}$) est de :..... 0.5 (2000/4000)

2°- Chaque item réussi rapporte $R(1-R) = 0.5(1-0.5) = + 0.25$

3°- Chaque item manqué coûte R^2..... $(0.5)^2 = - 0.25$

4°- L'écart entre "gain" et "perte" (AMPLITUDE) est de :
$R(1-R) + R^2 = R - R^2 + R^2 = R =$..... 0.5

5°- La somme des réussites est de : $(0.5)40 =$.....20
soit Qe x $\bar{R}$ ou ΣR

6°- La somme des pertes (Négatif maximum):
$\Sigma R^2 = - 40(0.5)^2 =$ ..... - 10

7°- La somme des gains (Positif Maximum):
$\Sigma[R(1-R)] = 40 \times 0.5 \times 0.5 = + 10$

Synthèses :

1°- $(\Sigma R^2) = [\Sigma R(1-R)]$ (Négatif)=(Positif).....(10)=(10)

2°- $\Sigma R = (\Sigma R^2) + (\Sigma R(1-R))$.....20 = (10) + (10)

3°- ΣR = Qe.$\bar{R}$ (A pour amplitude).....20 = 40 x 0.5

Remarque :

Le nombre de candidats a été négligé. En effet:

1°- $\frac{\Sigma R'}{C} = \bar{R}$: $\bar{R}$ est le rapport nombre d'item réussis au total sur le nombre de candidats. Nombre rapporté au nombre de candidats.

2°- Le correcteur a donné 2000 signes de satisfaction (+)soit 20 par candidat. Soit, pour chaque candidat, une fois pour deux item ou Qe x $\bar{R}$.

DEUXIEME EXEMPLE

Données :

. Qe = 40

. C = 100

. $\Sigma R'$ = 2000

. 20 item ont été réussis à 60 p.100

. 20 item ont été réussis à 40 p.100

Déductions :

1°- $\overline{R} = 2000/100 \times 40 = 2000/4000$ ..... $\overline{R} = 0.5$

2°- Certains item (à R = 0.6) rapportent :
$0.6(1-0.6) = 0.6 \times 0.4 =$ .....+0.24

3°- Certains item (à R = 0.4) rapportent :
$0.4(1-0.4) = 0.4 \times 0.6 =$ .....+0.24

4°- Certains item manqués (R = 0.6) coûtent $-(0.6)^2 = -0.36$

5°- Certains item manqués (R = 0.4) coûtent $-(0.4)^2 = -0.16$

6°- $\Sigma R = 20(0.6) + 20(0.4) = 12+8 = 20$ .....$=\overline{R}.Qe$....20

7°- $R^2 = 20(0.6)^2 + 20(0.4)^2 = 7.2 + 3.4 =$ 10.4

8°- $\Sigma R(1-R) = 20 \times 0.6(1-0.6) + 20 \times 0.4(1-0.4)$
$= 20 \times 0.6 \times 0.4 + 20 \times 0.4 \times 0.6 =$.....9.6

Synthèse :

1°- $\Sigma R = \overline{R}.Qe$ ...(ou A)

2°- $\Sigma R = (\Sigma R^2) + |\Sigma R(1-R)|$

3°- $|\Sigma R^2| \neq |\Sigma R(1-R)|$.....$|10.4| \neq |9.6|$
$|\Sigma R^2| > |\Sigma R(1-R)|$.....$|10.4| > |9.6|$

Remarques :

1°- Si les item (0.6) et (0.4) rapportent autant, leur coût d'échec est différent.

2°- L'écart entre "gain" et "perte" est ici $\Sigma R + |\Sigma R^2|$ soit 10.4 + 9.6 = 20 comme dans le premier exemple. On peut encore exprimer cet écart ou A (pour amplitude) par $A = \overline{R}Qe$ (0.5 x 40)

TROISIEME EXEMPLE

Données :

. Qe = 40

. C = 100

. $\Sigma R' = 2000$

. 10 item ont été réussis à 70 p.100 (0.7)

. 10 item ont été réussis à 60 p.100 (0.6)

. 10 item ont été réussis à 40 p.100 (0.4)

. 10 item ont été réussis à 30 p.100 (0.3)

Déductions :

1°- $\overline{R} = 2000/4000 = \ldots\ldots\ldots\ldots\ldots\ldots 0.5$

2°- $\Sigma R = 10(0.7) + 10(0.6) + 10(0.4) + 10(0.3) = 20 = \overline{R}.Qe$

3°- $\Sigma R^2 = 10(0.7)^2 + 10(0.6)^2 + 10(0.4)^2 + 10(0.3)^2$
$= 4.9 + 3.6 + 1.6 + 0.9 = \ldots\ldots\ldots\ldots\ldots 11 = N$

4°- $\Sigma R(1-R) = 10 \times 0.7(1-0.7)$
$+ 10 \times 0.6(1-0.6)$
$+ 10 \times 0.4(1-0.4)$
$+ 10 \times 0.3(1-0.3)$
$= 10 \times 0.7 \times 0.3 + 10 \times 0.6 \times 0.4$
$+ 10 \times 0.4 \times 0.6 + 10 \times 0.3 \times 0.7$
$= 2.1 + 2.4 + 2.4 + 2.1$
$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 9 = P$

Synthèse :

1°- $\Sigma R = \overline{R}.Qe$ (ou A)

2°- $\Sigma R = |\Sigma R^2| + |\Sigma R(1-R)|$

3°- $|\Sigma R^2| \neq |\Sigma R(1-R)| \ldots\ldots\ldots\ldots\ldots\ldots |11| \neq |9|$

4°- $|\Sigma R^2| > |\Sigma R(1-R)| \ldots\ldots\ldots\ldots\ldots\ldots |11| > |9|$

QUATRIEME EXEMPLE -

Données :

. Qe = 50

. C = 100

. $\Sigma R' = 2500$

. 10 item ont été réussis à 70 p.100 (R = 0.7)

. 10 item ont été réussis à 60 p.100 (R = 0.6)

. 10 item ont été réussis à 50 p.100 (R = 0.5)

. 10 item ont été réussis à 40 p.100 (R = 0.4)

. 10 item ont été réussis à 30 p.100 (R = 0.3)

Déductions :

1°- $\overline{R} = 2500/50 \times 100 = 2500/5000$ .....= 0.5

2°- $\Sigma R = 10(0.7) + 10(0,6) + 10(0.5) + 10(0.4) + 10(0.3) =$
$\overline{R}.Qe = 25$

3°- $\Sigma R^2 = 10(0.7)^2 + 10(0.6)^2 + 10(0.5)^2 + 10(0.4)^2 + 10(0.3)^2 =$
N =-13.5

4°- $\Sigma R(1-R) = 10 \times 0,7 \times 0.3 +$
$+10 \times 0.6 \times 0.4 +$
$+10 \times 0.5 \times 0.5 +$
$+10 \times 0.4 \times 0.6 +$
$+10 \times 0.3 \times 0.7$
$= 2.1 + 2.4 + 2.5 + 2.4 + 2.1$ .....=11.5
=P

Synthèse :

1°- $\Sigma R = \overline{R}.Qe$ (ou A)

2°- $\Sigma R = |\Sigma R^2| + |\Sigma R(1-R)|$

3°- $|\Sigma R^2| \neq |\Sigma R(1-R)|$ ..... $|13.5| \neq |11.5|$

4°- $|\Sigma R^2| > |\Sigma R(1-R)|$ ..... $|13.5| > |11.5|$

CINQUIEME EXEMPLE

Données

. Qe = 40

. C = 100

. $\Sigma R' = 2000$

. 20 item ont été réussis à 90 p.100 (R = 0.9)

. 20 item ont été réussis à 10 p.100 (R = 0.1)

Déductions :

1°- $\overline{R} = 2000/4000 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 0.5$

2°- $\Sigma R = 20 \times 0.9 + 20 \times 0.1 = 18 + 2 = \ldots\ldots\ldots\ldots 20 = \overline{R}.Q$

3°- $\Sigma R^2 = 20 \times (0.9)^2 + 20 \times (0.1)^2$
$= 20 \times 0.81 + 20 \times 0.01 = 16.2 + 0.2 = \ldots\ldots\ldots 16.4 = N$

4°- $\Sigma R(1-R) = 20 \times 0.9 \times 0.1 + 20 \times 0.1 \times 0.9$
$= 20 \times 0.09 + 20 \times 0.09 = 1.8 + 1.8 = \ldots 3.6 = P$

Synthèse :

1°- $\Sigma R = \overline{R}.Qe = A$

2°- $\Sigma R = |\Sigma R^2| + |\Sigma R(1-R)|$

3°- $|\Sigma R^2| \neq |\Sigma R(1-R)| \ldots\ldots\ldots\ldots\ldots\ldots |16.4| \neq |3.6|$

4°- $|\Sigma R^2| > |\Sigma R(1-R)| \ldots\ldots\ldots\ldots\ldots\ldots |16.4| > |3.6|$

RESUME

Tous les exemples ont un point commum : $\overline{R} = 0.5$

	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5
Qe	40	40	40	50	40
Qe.C	4000	4000	4000	5000	4000
$\Sigma R'$	2000	2000	2000	2500	2000
$\overline{R}$	0.5	0.5	0.5	0.5	0.5
ΣR	$Qe.\overline{R}$	$Qe.\overline{R}$	$Qe.\overline{R}$	$Qe.\overline{R}$	$Qe.\overline{R}$
ΣR^2 ou N	10	10.4	11	13.5	16.4
$\Sigma R(1-R)$ ou P	10	9.6	9	11.5	3.6
ΣR ou A	20	20	20	25	20

Un certain nombre d'aspects communs se dégagent :

1°- $\Sigma R = Qe.\overline{R}$

2°- $\Sigma R = |\Sigma R^2| + |\Sigma R(1-R)|$

Malgré ces points communs quelques différences se manifesten

1°- Sauf pour le 1er exemple: $|\Sigma R^2| \neq |\Sigma R(1-R)|$

2°- Si cette inégalité se trouve elle est toujours dans le même sens:

$$|\Sigma R^2| > |\Sigma R(1-R)|$$

Ainsi quelques grandeurs intéressantes se trouvent démasquées.

QUELQUES GRANDEURS INTERESSANTES

Pour qu'une grandeur soit jugée intéressante, il faut qu'elle ait une signification.

Il s'agit ici d'approcher leur signification et non de la définir ce qui ne pourra être fait que dans la mesure où les hypothèses pressenties seront démontrées. Pour la clarté de ce qui suit, il a semblé nécessaire de représenter chacune de ces grandeurs de trois manières:

1) De la manière dont elles ont été obtenues

   . ΣR

   . $\Sigma R(1-R)$

   . ΣR^2

2) Par ce qu'elles représentent concrètement, respectivement

   . L'amplitude (de la dispersion)

   . Le positif maximum attribuable

   . Le négatif maximum attribué

3) Par des symboles propres à entrer dans des formules, respectivement:

   . A

   . P

   . N

1) ΣR ou A (A pour Amplitude)

Des exemples choisis, il ressort que cette valeur ne dépend, pour une valeur donnée de Qe, dont l'examinateur est maître, que de $\overline{R}$, qui est une grandeur à priori imprévisible. Il a toujours été trouvé:

$$\Sigma R = Qe.\overline{R}$$

Ceci est bien une loi générale: la somme des réussites à chaque item (ΣR) est identique à la réussite moyenne de chaque item ($\overline{R}$) multiplié par le nombre d'item (Qe).

Si on représente le nombre de réponses acceptées dans l'item 1 par R'_1

Si " " " " " " " " 2 " R'_2

Si " " " " " " " " n " R'_n

On a: $\frac{R'1}{C} + \frac{R'2}{C} + \ldots \frac{R'n}{C} = R_1 + R_2 + \ldots R_n = \Sigma R$

ou $R'_1 + R'_2 + \ldots R'_n = \Sigma R'$

$$\frac{\Sigma R'}{C} = \Sigma R$$

$$\frac{\Sigma R}{Qe} = \overline{R} \quad \text{donc} \quad \Sigma R = \overline{R}.Qe$$

ΣR est une grandeur qui représente en réalité la somme des valeurs attribuées à chaque item: elle représente le plus grand SPI possible à partir du zéro absolu (c'est-à-dire sans notation négative).

On peut tout aussi bien lui attribuer une autre fonction représentative :

. $\overline{R}$.Qe, est un nombre fictif d'item; que chaque item soit réussi à 50 p.100 ou que 50 p.100 des item soient complétement réussis (R = 1) et 50 p.10 complètement manqués (R = 0), la valeur de $\overline{R}$.Qe serait la même.

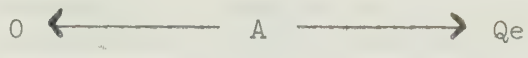
On peut donc dire que $\overline{R}$.Qe représente le nombre d'item réussis par candidat dans l'épreuve en supposant qu'ils sont également réussis par chacun: c'est le Score Pondéré Individuel Moyen (SPIM).

Cette grandeur, que nous appellerons l'Amplitude du résultat (A) vaudrait Qe si tous les item étaient complètement réussis ; elle vaudrait zéro si aucun item n'était réussi : si tous étaient complètement manqués.

Dans le premier cas, A serait de la forme Σ0 avec Σ équivalent de Qe puisque tous les item sont "réussis" à R = 0 .

Dans le second cas, A vaudrait Σ1, avec Σ équivalent de Qe pour la même raison.

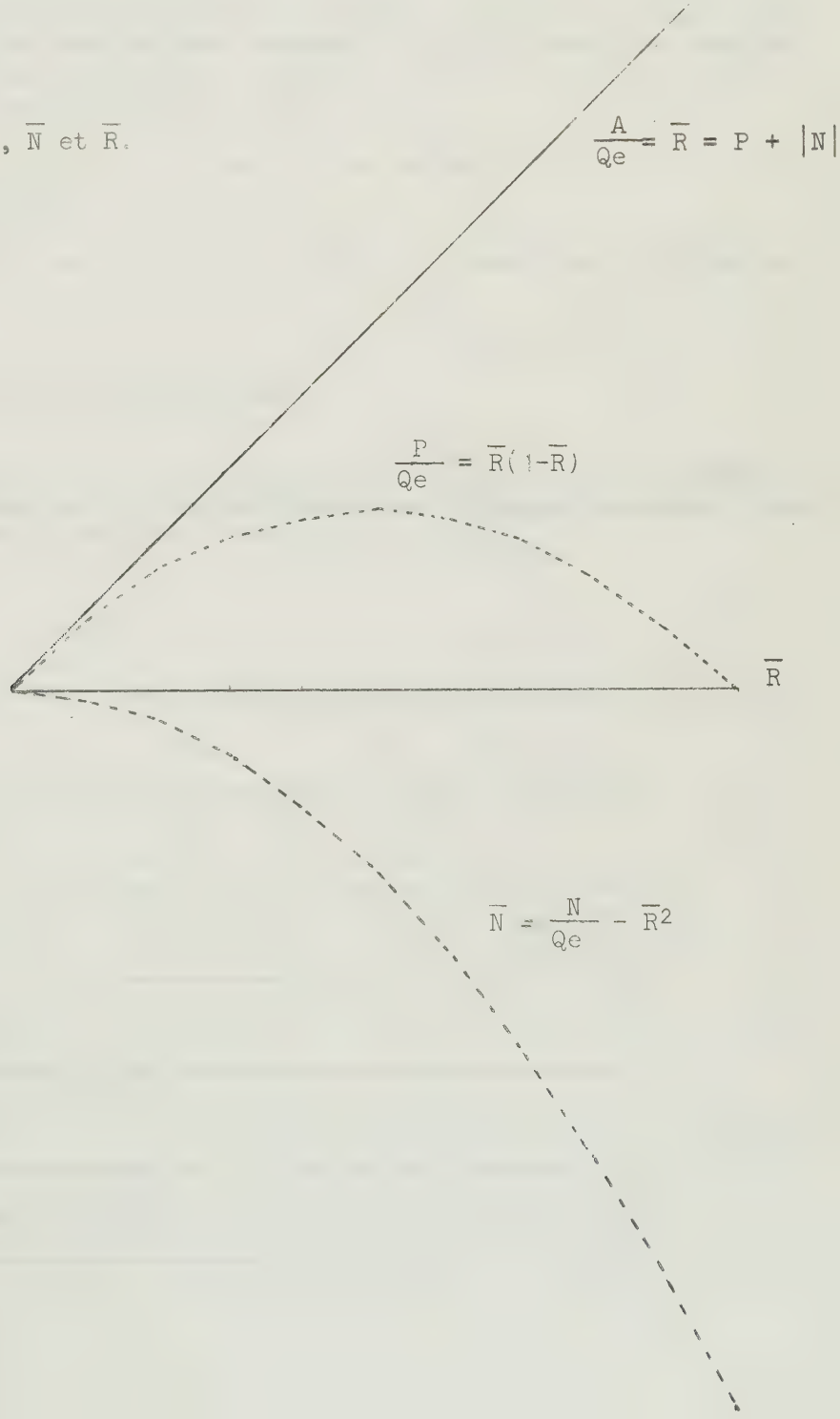
Ainsi A mesure la difficulté moyenne de l'examen . Cette grandeur connaît deux limites parfaitement définies :



Une autre signification peut lui être attribuée : A exprime le nombre d'item réussis par le "candidat moyen", à supposer que chaque item ait la même valeur . Dans les exemples précédents, on peut dire que le candidat moyen a réussi environ 50 p.100 des item . Mais cet "environ" méritera d'être précisé .

Relations entre $\overline{A}$, $\overline{P}$, $\overline{N}$ et $\overline{R}$.

(cf. p.241)



2) $\Sigma R(1-R)$ ou P (Positif)

Rapportée au candidat moyen, cette grandeur exprime le meilleur SPI qu'il puisse espérer. Compte tenu de la valeur différente attribuée à chaque item, le SPI ne peut pas dépasser cette valeur. Il exprime le positif maximum réalisable.

On peut le calculer de deux manières:

1/ Positif calculé (Pc) ou prévisible.

Il se calcule à partir de données schématiques. Dans les exemples choisis un résultat est constant: $\bar{R} = 0.5$. Dans les exemples 1, 2, 3 et 5, où Qe valait 40, on pouvait estimer P par calcul:

$P = Qe.R(1-R) = 40 \times 0.5(1-0.5) = 10.$

Dans l'exemple 4 (Qe = 50) P pouvait être estimé à:

$P = 50 \times 0.5 \times 0.5 = 12.5.$

On voit que ce P calculé (Pc) ne dépend que de $\bar{R}$ pour un Qe donné.

2/ Positif observé (Po).

Apartir des données observées on établit Po en faisant réellement la somme des R(1-R) de chaque item. Presque toujours, à une exception près, Po s'est montré plus petit que Pc:

	Po	Pc	Pc - Po
Ex.1	10	10	0
Ex.2	9.6	10	0.4
Ex.3	9	10	1
Ex.4	11.5	12.5	1
Ex.5	3.6	10	6.4

3) ΣR^2 ou N (Négatif)

Rapporté au candidat moyen c'est le négatif maximum qu'il doit éviter. C'est le handicap. Il est possible de calculer N, comme P, à partir de deux sortes de données:

1/ Négatif calculé (Nc) ou prévisible.

Il se calcule à partir des données schématiques:

$$\Sigma R^2 = Qe.\overline{R}^2.$$

Dans les exemples précédents (1, 2, 3 et 5) ce négatif calculé (Nc) valait: $40 \times (0.5)^2 = 10$.

Dans l'exemple 5, Nc valait:
$50 \times (0.5)^2 = 12.5$.

2/ Négatif observé (No).

A partir des données observées, on établit No en faisant réellement la somme des R^2. Voici ce qu'on obtenait pour chacun des ex-exemples:

	No	Nc	No - Nc	Pc - Po (rappel)
Ex. 1	10	10	0	0
Ex. 2	10.4	10	0.4	0.4
Ex. 3	11	10	1	1
Ex. 4	13.5	12.5	1	1
Ex. 5	16.4	10	6.4	6.4

On peut remarquer:

a) que tous les négatifs observés, sont, en valeur absolue, supérieurs aux négatifs calculés, sauf dans le premier exemple.

b) que les écarts No-Nc sont partout égaux aux différences Pc-Po.

SIGNIFICATION DE CES DIFFERENCES

1/ Egalité de (Pc-Po) et (No-Nc)

Elle n'a rien d'étonnant. De toutes façons, P et N sont liés à A selon la relation: $A = |P| + |N|$. Or, A ne dépend que de Qe et de $\bar{R}$.

Dans les exemples 1, 2, 3 et 5, A est constant et s'obtient ainsi ($A = \Sigma R$ de la forme R).

$A = \Sigma R(1-R)$ ce qui est identique à:
$A = \Sigma (R-R^2)$ ou à
$A = \Sigma R - \Sigma R^2$

On peut tirer de la dernière expression:

$\Sigma R = A - |\Sigma R^2|$
$\Sigma R^2 = A - |\Sigma R|$ et pour A constant
$\Sigma R + \varepsilon = A - |\Sigma R^2| - \varepsilon$.

Pour une épreuve donnée, caractérisée par un $\bar{R}$ et une valeur de Qe, la somme de ces deux termes étant constante, ce qui est ajouté à l'un des deux est retranché de l'autre.

En pratique, il est plus simple de déduire P des autres valeurs:

$P = A - |N|$ ou $\Sigma R(1-R) = \Sigma R - \Sigma R^2$.

Sachant que les différences (Pc - Po) et (No - Nc) sont nécessairement égales il reste à expliquer pourquoi une différence entre No et Nc.

2/ Comparaisons entre Po et (No - Nc)

L'observation du tableau de la page suivante montre que:

1) A la plus grande valeur de Po correspond la plus petite valeur de (No - Nc) et inversement, entre 1 et 5 quelque chose diffère beaucoup.

2) Que, malgré des valeurs différentes de Po en 3 et 4, leur (No - Nc) sont les mêmes: entre 3 et 4, quelque chose est comparable.

Sans savoir si l'exemple 1 présente le cas du plus grand Po possible, il est, de la liste, le plus grand observé. Quant à la différence No-Nc elle semble le plus petit possible, et, en valeur absolue

c'est certainement la plus petite possible. Comment se présente donc l'exemple 1 ?

1) Tous les item ont été réussis à 50 p.100; chacun d'entre eux à 50 p.100. C'est monotone: ce n'est pas varié.

2) On ne sait rien des candidats mais, manifestement, le nombre de combinaisons qui leur est offert pour réussir est grand. Le plus grand possible, puisqu'ils peuvent réussir n'importe quel item , la seule condition qui leur soit imposée c'est d'en réussir 50 p.100.

Au total, il semble bien que ces grandeurs soient en rapport avec la variété.

	Po	No - Nc
Ex. 1	10	0
Ex. 2	9.6	0.4
Ex. 3	9	1
Ex. 4	11.5	1
Ex. 5	3.6	6.4

Considérons l'exemple 5: Po y est beaucoup plus petit et No - Nc beaucoup plus grand qu'en 1.

1) Certains item ont été réussis à 10 p.100 d'autres à 90 p.100: c'est plus varié.

2) Les candidats n'ont pas un très grand nombre de possibilités pour être reçus. Comment peut-on être reçu en 5 ?

a) avec un minimum de réussite si on choisit les plus "chères", (0.9). Le handicap est de -16.4 ($-\Sigma R^2$), il faut donc réaliser un R au moins égal à $|\Sigma R^2|$.

.1 item à 0.9 donne un SPI de:..... + 0.9-16.4
.18 item à 0.9 donnent un SPI de: ..... + 16.2-16.4
.19 item à 0.9 donnent un SPI de: ..... + 17.1-16.4:RECU

b) avec un maximum de réussite si on choisit les "moins chères" (0.1)

.1 item à 0.1 donne un SPI de: + 0.1-16.4
.20 item à 0.1 donnent un SPI de: + 2 -16.4
.20 item à 0.1 + 10 item à 0.9 donnent un SPI de:
+ 2 + 9 = + 11 -16.4

. 20 item à 0.1 + 15 item à 0.9 donnent un SPI de:
+ 2 + 13.5 = + 15.5 - 16.4
. 20 item à 0.1 + 16 item à 0.9 donnent un SPI de:
+ 2 + 14.4 = 16.4 - 16.4:REC

Si on "choisit" bien, il suffit de réussir 19 item sur 40. Si on choisit mal, il en faut 36.

Au total, il semble que ces grandeurs soient liés à la grandeur de la variété.

La comparaison des exemples 3 et 4 va préciser la signification. Voici deux examens qui ont même réussite moyenne (comme tous les autres: $\bar{R} = 0.5$) qui n'ont pas le même nombre d'item, qui n'ont pas la même valeur de Po mais qui ont la même valeur de No - Nc. Qu'ont donc de semblable ces deux examens et en quoi diffèrent-ils ?

Ils diffèrent en ceci: le nombre d'item de 4 est supérieur à celui de 3. La valeur de Po en 4 est supérieure à celle de 3: rien d'étonnant à ce que Po varie dans le même sens que le nombre d'item, puisque le SPI MAXIMUM possible dépend, toutes choses égales par ailleurs, du nombre d'item et Po représente notamment ce SPI maximum possible.

Qu'ont-ils de commun ? La distribution de la réussite des item. Comparons l'exemple 3 (ligne supérieure) et l'exemple 4 (ligne inférieure).

Nombre d'item réussi à	0.7	0.6	0.5	0.4	0.3
Exemple 3	10	10	-	10	10
Exemple 4	10	10	10	10	10

Dans l'un comme dans l'autre cas, 20 item s'écartent de la moyenne de ± 0.1 et 20 item de ± 0.2.

Au total, il semble bien que ces deux grandeurs soient des VARIANCES.

DES VARIANCES OBTENUES PAR SOUSTRACTIONS !

Po qui se calcule commodément par $\Sigma R - \Sigma R^2$, vaut $\Sigma R(1-R)$. $\Sigma R(1-R)$ celà peut s'écrire: $Qe.\bar{R} - \Sigma R^2$ puisque $\Sigma R = Qe.\bar{R}$.

Ceci peut encore se représenter par Qe.R.M (M pour manqué) puisque tout ce qui n'est pas réussi est manqué (jugement binaire) et que R + M = 1. Tout ce qui n'est pas réussi est manqué: ceci n'est pas une tautologie: il faut souligner que ce mode de correction exige une certitude: "accepté" ou "refusé", sans nuance, ni intermédiaire. Notamment un douteux est exclu.

On a bien: $R + M = 1$ avec $M = 1 - R$, donc :
$$R + (1 - R) = R + 1 - R = 1.$$

Par conséquent, Po peut s'écrire sous la forme Qe.R.M.identique à nqp où n représente un nombre (Qe = nombre d'item); p: la proportion des réussites (R) et q la proportion d'échecs (M) avec p + q = 1: et R + (1 - R) = 1.

Ceci est une variance exprimée, en nombre (et non en pourcentage).

No - Nc est obtenu par une soustraction de la forme:

$$\Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

No est donné par ΣR^2

Nc par $Qe.\bar{R}^2$ or, $Qe.\bar{R}^2$ est différent de $(Qe.\bar{R})^2$

$(Qe.\bar{R})^2$ développé donne $Qe^2\bar{R}^2$ donc $Qe.\bar{R}^2 = \frac{(Qe.\bar{R})^2}{Qe}$

Ainsi $\Sigma R^2 - Qe.\bar{R}^2$ est bien de la forme: $\Sigma x^2 - \frac{(\Sigma x)^2}{n}$ et cette forme est celle d'un numérateur de variance:

$$S^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}$$

QUE REPRESENTENT CES VARIANCES ?

Il est visible qu'elles ne représentent pas la même chose et, d'abord parce qu'elles évoluent selon les circonstances, en sens inverse l'une de l'autre.

Po, autrement dit le positif observé ou $\Sigma R(1-R)$ est la représentation du SPI maximum d'une part, et la variance de cette réussite d'autre part.

$\sqrt{\Sigma R(1-R)}$ (qui a bien la forme $\sqrt{npq}$) représente l'écart-type exprimé en item réussis en supposant non pas qu'on additionne des item mais leurs valeurs. Po représente un nombre réduit d'item, nombre exprimé par la somme des valeurs qui est attribuée à chacun d'eux.

$\sqrt{Po}$ représentant l'écart-type, $\pm 2\sqrt{Po}$ représentent les valeurs extrêmes du SPI pour 95p.100 de la population en supposant qu'elle soit normale. Mais tous les calculs ayant été faits sur le nombre d'item et non sur celui des candidats, il importe peu que la population des candidats soit normale. Au contraire, ce test permet d'en juger.

Ainsi, et pour reprendre l'exemple 3, $\Sigma R(1-R)$ valait 9. L'écart-type est donc 3 et deux écarts-types valent 6. Une population normale de candidats doit se grouper à 95 p.100 entre - 6 et + 6.

En ce qui concerne No - Nc, l'exemple (1) mettra sur la voie Dans cet exemple, No - Nc vaut 0: la variance est nulle. Les données indiquent immédiatement ce qui ne varie pas: la réussite des item. Tous les item ont été réussis à 50 p.100 (dispersion nulle). Ainsi cette variance est celle de la réussite des item.

De l'exemple (3) où No - Nc vaut 1, on tire la variance:

$S^2 = \frac{1}{40 - 1}$ enprenant $\frac{1}{40 - 1}$ sensiblement égal à $\frac{1}{40}$, on tire:

$S^2 = 0.025000$. L'écart-type ($\sqrt{S^2}$ ou S) vaut 0.158. Deux écarts-types valent 0.316. Ce qui signifie, qu'au risque 5 p.100, 95 p.100 des item se trouvent avoir un coefficient de réussite valant $\bar{R} \pm 2S$ soit ici $0.184 \longleftarrow 0.5 \longrightarrow 0.816$. En fait, la distribution réelle, concernant la totalité des item, est $0.3 \longleftarrow 0.5 \longrightarrow 0.7$: la distribution n'est pas normale et cette constatation n'a rien d'étonnant puisque l'histogramme de (3) s'inscrit dans un rectangle et non pas dans une courbe de Gauss. Il ne faut pas oublier que tous les exemples cités sont fictifs.

SIGNIFICATION DES GRANDEURS A, P, N, $\overline{R}$

A -

A = P + |N|. C'est l'amplitude de la dispersion qui est limitée par :

- un minimum : $-\Sigma R^2$: aucune réponse acceptée ou feuille blanche.
- un maximum : $+\Sigma R(1-R)$: toutes les réponses ont été acceptées.

P -

Po, celui qu'on observe, est le SPI maximal que puisse espérer un candidat. Cette grandeur est de la forme d'une variance : dispersion autour de la moyenne qui est, ici, zéro. Un candidat "normal" n'a que cinq chances sur cent de s'écarter de la moyenne de plus de $2\sqrt{Po}$ dans un sens ou dans l'autre. Po représente donc une "espérance" maximum.

N -

No, celui qu'on observe, est le handicap, c'est le SPI minimal que doit atteindre le candidat pour avoir zéro c'est-à-dire être reçu : il faut que ce SPI soit au moins égal à ΣR^2 puisque on est reçu qu'avec, au moins, |P| - |N| = 0.

La différence No - Nc avec |No| > |Nc| représente la dispersion des réussites des item autour d'une moyenne qui vaut $\overline{R}$.

$\overline{R}$ -

C'est la réussite moyenne des item. On peut l'obtenir soit en rapportant le total des item acceptés à celui des item corrigés :

$$\frac{R'}{Qe \times C}$$

Soit en faisant la moyenne des réussites de chaque item :

$$\frac{R_1 + R_2 + \ldots . R_n}{Qe} = \frac{\Sigma R}{Qe}$$

Soit en établissant le rapport $\frac{A}{Qe}$ ce qui revient à $\frac{\Sigma R}{Qe}$.

Le tableau suivant indique les rapports qui unissent ces grandeurs. Rappelons que Ac = Ao mais que Po ≠ Pc et No ≠ Nc avec |Po - Pc| = |No - Nc|.

	A/Qe	Pc/Qe	- Nc/Qe
$\bar{R} = 0$	0	0	- 0
$\bar{R} = 0.1$	0.1	0.09	- 0.01
$\bar{R} = 0.2$	0.2	0.16	- 0.04
$\bar{R} = 0.3$	0.3	0.21	- 0.09
$\bar{R} = 0.4$	0.4	0.24	- 0.16
$\bar{R} = 0.5$	0.5	0.25	- 0.25
$\bar{R} = 0.6$	0.6	0.24	- 0.36
$\bar{R} = 0.7$	0.7	0.21	- 0.49
$\bar{R} = 0.8$	0.8	0.16	- 0.64
$\bar{R} = 0.9$	0.9	0.09	- 0.81
$\bar{R} = 1.0$	1.0	0.00	- 1.00

Plusieurs constatations s'imposent:

1/ $|A/Qe(\bar{A})|$ reflète la difficulté moyenne de l'épreuve: en fait c'est l'expression de sa facilité.

2/ $Pc/Qe(\bar{Pc})$ croit pour des $\bar{R}$ allant de 0 à 0.5 et décroît ensuite. Ces variations représentent la loi de probalité de $\bar{R}$ de forme: variation de $\bar{R}.\bar{M} = \bar{R}(1-\bar{R})$ en fonction de $\bar{R}$. Et c'est bien ainsi qu'elle avait été exprimée. Plus $\bar{Pc}$ est élevé, plus l'épreuve est "moyenne". Vers les extrêmes, c'est $\bar{A}$ qui indique la facilité de l'épreuve.

3/ Nc/Qe ($\bar{Nc}$) croît avec la facilité de l'épreuve: plus l'épreuve est facile, plus le handicap est lourd.

4/ Si on se rappelle que $|No| > |Nc|$, que $|No| - |Nc|$ exprime la variété dans la réussite des item, que plus cette variété est grande, plus No est grand, on s'aperçoit que la variété d'une épreuve est facteur de difficulté qui augmente No (donc le handicap), et diminue l'espérance de réussite (Po).

Ainsi, ce système, fondé en fait sur le "résumé de l'épreuve" : sa moyenne et sa variance, est réaliste. Il permet d'obtenir des notes significatives.

ET LES CANDIDATS ?

Pris un par un, chacun est ce qu'il est. Pris dans l'ensemble ils sont probablement variés, c'est-à-dire qu'ils s'éloignent plus ou moins, dans un sens ou dans l'autre, des normes du candidat moyen.

Un exemple très simple (et impropre aux calculs car les nombres sont trop petits): une épreuve de deux item subie par deux candidats. Chaque item est réussi une fois sur deux. Celà peut aussi bien vouloir dire que chaque item a été réussi chacun par un candidat, ou que les deux item ont été réussis par le même: la réussite de chacun d'eux ne sera évidemment pas la même dans les deux cas.

Soit une épreuve ainsi résumée:

$$Qe = 40 \; ; \; C = 100 \; ; \; \Sigma R = 40 \; .$$

$$\overline{R} = \frac{40}{40 \times 100} = \frac{40}{4000} = 0.01.$$

Voici un examen peu réussi mais on ne peut savoir avec la moindre précision combien de candidats sont reçus.

- Ou bien les 40 item réussis l'ont été par le même candidat: il a répondu juste partout, il est reçu. Tous les autres sont refusés avec un SPI négatif maximal.

- Ou bien chaque item a été réussi par un candidat différent: chacun de ces 40 candidats est reçu.

- Ou bien un seul item a été réussi par 40 candidats : chacun de ces 40 candidats est reçu.

Le nombre des reçus est compris entre 1 et 40 sur 100: l'effectif reçu et la moyenne de l'épreuve sont largement indépendants.

Il y a au moins un candidat reçu: il a répondu juste partout et il est normal qu'il soit reçu. Par ailleurs, puisque des réponses justes ont été faites, il fallait bien qu'elles le fussent par quelqu'un.

Il peut paraître choquant qu'avec aussi peu de succés dans les réponses, 40 p.100 des candidats puissent être reçus. A celà on peut rétorquer que cette épreuve a donné ces résultats sur certains item et que ce ne sont pas les candidats qui les ont choisis.

Une évaluation sensible et fidèle et même juste ne suffit pas pour qu'une épreuve soit significative.

CHAPITRE IV

UNE NOTATION NORMALISEE

B - DONNER UNE SIGNIFICATION AUX NOTES

PRINCIPES DE LA NOTATION 244
QUELLES GRANDEURS CHOISIR? 245
L'ECART TYPE ENVISAGE COMME UNITE DE NOTATION 246
. Comment l'obtenir 246
. Définition des notes extrêmes: 0/20 et 20/20 246
. Critique de ces définitions 246
INTERVALLE DE CONFIANCE DE LA MOYENNE ET "NOTES MOYENNES" 248
. Comment obtenir l'écart-type de la moyenne 248
. Définition du neuf 248
. Limite supérieure du ONZE 248
. Frontière entre DIX ET ONZE 249
. Définition des " notes moyennes " 249
. Critique de ces définitions 249
DEFINITION DES AUTRES NOTES 250
. Une frontière mal définie: 11-12 (ou 8-9) 250
. Rapport S/S' 250
. Un système séduisant 251
. Une zone difficile à définir: ONZE^{+} DOUZE^{-} 251
DEFINITION DES 21 DEGRES ou DE 20 INTERVALLES 253
REPARTITION D'UNE POPULATION NORMALE 254
COMPENSATIONS OU ROLE DU JURY 258
NOTATION ET COEFFICIENT NORMATIF 262
CONCLUSION 263

Le calcul d'une note normalisée est assez complexe. Le résultat obtenu par les calculs précédents (chapitre II), suffit non seulement à indiquer si la valeur du candidat atteint les normes fixées, mais encore à établir un classement dans une composition ou un concours même organisé en vue d'un "Numerus clausus". Si on souhaite recruter un nombre restreint de candidats, le simple classement suffit et comme il a déjà été largement vu, le nombre d'item et la précision des calculs permettent une discrimination pratiquement illimitée. Si on souhaite que les candidats recrutés soient, en plus, suffisants, le zéro peut servir de seuil même si des places restent à pourvoir.

Utiles ou non, les notes d'examen sont souvent justifiées par la prescription réglementaire et le culte que leur portent les candidats et leurs familles: ces trois facteurs ne semblant pas parfaitement indépendants. Si on doit en mettre, autant que ce soit d'une façon telle qu'elles soient significatives.

La note, s'il en faut une, doit être représentative de l'éloignement de chacun des candidats par rapport au candidat "moyen" de résultat ZERO. L'unité choisie a été l'écart type de la distribution supposée normale des Scores Pondérés Individuels par rapport à la "moyenne" c'est à dire à la norme minimale de suffisance. Ces deux propositions sont arbitraires mais non fantaisistes.

PRINCIPES DE LA NOTATION

OBJECTIFS DE LA NOTATION.

1°- permettre des comparaisons

2°- attribuer une note telle,qu'il soit certain que 12 représente une valeur supérieure à celle qu'exprime 10.

La notation n'a pas pour objectif d'affirmer que 12 représente une valeur double de six.

PREMIERE DEMARCHE:

Choisir une échelle d'expression: un code, un langage communs à l'intérieur d'une même discipline et d'une discipline à l'autre si l'examen est composé de plusieurs épreuves cumulatives. L'échelle est souvent imposée par voie réglementaire: notation de 0 à 20. Il faut que chacun des échelons soit défini.

DEUXIEME DEMARCHE:

Si les épreuves d'un même examen sont cumulatives, choisir la limite de compensation d'une épreuve par une autre. C'est dire que si les réussites peuvent être très différentes d'une épreuve à l'autre, il est absolument nécessaire que la note soit mise selon des règles communes rigoureusement définies.

QUELLES GRANDEURS CHOISIR ?

Evidemment des grandeurs objectives, tirées du score de chaque épreuve.

D'une épreuve à l'autre dans une même discipline, et, à fortiori, d'une discipline à l'autre, les scores peuvent être très différents. Ce n'est pas une nécessité mais une possibilité. Constater ce fait peut inciter à des réflexions pédagogiques mais ne doit en aucun cas mettre en cause la méthode d'établissement du score. Une discipline (ou un correcteur) peut manifester sa personnalité par la valeur de la moyenne et la dispersion autour de cette moyenne dans le système traditionnel. Dans le système de cotation proposé, la moyenne est pour tout le monde égale: c'est zéro. C'est une grandeur parfaitement définie qui s'exprime parfaitement par 10 sur 20.

En admettant que 10 sur 20 représente la plus petite note requise pour être reçu, un S.P.I."zéro" se situe dans le onzième degré de l'échelle vigintésimale si zéro, de son côté, représente indiscutablement le plus petit score pondéré satisfaisant. Est-ce le cas ?

L'exemple (5) a montré que pour avoir "zéro de réussite" un candidat devait réussir:

$$\begin{array}{r} 18 \text{ item à } 0.9 \ldots + \ 16.2 \\ + \ 2 \text{ item à } 0.1 \ldots + \ \underline{0.2} \\ \hline 16.4 \end{array} - 16.4\ (\Sigma R^2) = \text{ZERO}.$$

Ainsi il devait non seulement réussir 20 item sur 40 (20 sur 40 n'étant, ici, qu'une simple coïncidence), mais, parmi ces 20, neuf fois plus d'importants que de moins importants, c'est-à-dire un nombre directement proportionnel à l'importance attachée à l'item.

Même dans cet exemple, où il était possible d'être reçu avec seulement 19 sur 40 item (à condition de les bien choisir), une réussite de 36 sur 40 pouvait être exigée pour être juste "moyen" dans une épreuve dont le coefficient de réussite était pourtant de 50p.100.

Ainsi, la moyenne étant strictement définie et plus sur des critères qualitatifs que purement quantitatifs, son caractère "satisfaisant" peut être retenu.

L'autre grandeur est la dispersion, exprimée en écart type. Elle s'impose, mais son maniement exige qu'on en connaisse les conséquences. Alors, et alors seulement,on pourra parler de choix.

L'ECART TYPE ENVISAGE COMME UNITE DE NOTATION

COMMENT L'OBTENIR ?

C'EST LA RACINE DE Po. Son calcul est très simple et aucun enseignant ne peut être arrêté par la nécessité d'utiliser une table. On le représentera par S:

$$S = \sqrt{Po}$$

DEFINITION DES NOTES EXTREMES: 0 SUR 20 ET 20 SUR 20.

Dans une distribution gaussienne, par conséquent concernant une population distribuée au hasard autour d'une moyenne, 95p.100 (environ) des individus sont groupés, de part et d'autre de la moyenne sans en être éloignés, dans un sens ou dans l'autre, de plus de deux écarts types. Le corollaire de cette proposition, c'est que les individus éloignés de la moyenne de plus de deux écarts types, n'appartiennent pas, statistiquement parlant, à la population étudiée. Ceci peut se manifester clairement par deux notes: ZERO et VINGT.

CRITIQUE DE CES DEFINITIONS.

Ne pas appartenir à la population est défini par un éloignement de la moyenne constaté supérieur à deux écarts types. Non seulement cette définition est arbitraire mais elle est fausse ou, plus exactement, elle n'est pas certainement vraie. Le critère choisi comporte un risque d'erreur connu: 5p.100.

Il eût été plus rigoureux de prendre trois écarts types et le risque d'erreur ne serait plus que de 0.3p.100. En augmentant le risque, on augmente aussi la probabilité d'attribuer un vingt sur vingt ce qui semble psychologiquement important, cette note retentissante n'étant réputée inaccessible que si personne n'y accède. Il en est de même du zéro sur vingt qui ne semble pas devoir être un épouvantail mais une réalité accessible. Ces deux notes n'auraient pratiquement aucune chance d'être attribuées si la référence était la variance ce qui lui fait préférer l'écart type malgré la grande difficulté psychologique que représente, peut être, pour certains enseignants, l'extraction d'une racine carrée.

Il importe de souligner que ces grandeurs sont celles qui caractérisent une population gaussienne, prise comme référence. Celà signifie qu'il est supposé que les individus sont répartis au hasard autour d'une moyenne. La constatation d'une répartition non gaussienne indique simplement que la population est organisée. Il est alors tout à fait possible que tous les candidats, ou presque, aient vingt sur vingt ou

zéro sur vingt selon leur sens de l'organisation.

Au total:

. Définition du ZERO des notes: SPI < m - 2S (candidats exceptionnellement mauvais)

. " du VINGT " : SPI > m + 2S (candidats exceptionnellement bons)

Et se trouvent approchées trois autres définitions:

. UN des notes : SPI de l'ordre de m - 2S

. DIX-NEUF " : SPI " " " m + 2S

. DIX " " : SPI " " " m (SPI = zéro)

INTERVALLE DE CONFIANCE DE LA MOYENNE ET "NOTES MOYENNES"

L'échelle des notes étant définie par des intervalles et non par des "points", il n'est pas possible, raisonnablement de définir l'intervalle représenté par dix sur vingt par le seul ZERO du Score Pondéré Individuel puisque ce ZERO est un point. La question est alors de définir, par ses deux extrémités, cet intervalle.

COMMENT OBTENIR L'ECART TYPE DE LA MOYENNE.

A partir du score, évidemment. Cet écart type qui:

. se représente par S'

. vaut $\frac{S}{\sqrt{Qe}}$

. S n'est pas fourni directement par le score mais sa valeur est facilement obtenue par un renseignement que fourni le score: Po dont S est la racine carrée.

. $S' = \frac{\sqrt{Po}}{\sqrt{Qe}}$, Po et Qe étant fournis par le score.

DEFINITION DU NEUF DES NOTES.

. Partant du principe que dix sur vingt désigne un SPI valant au moins ZERO et, par conséquent un candidat qui s'est montré au moins moyen,

. le neuf désignera un candidat inférieur à la "moyenne" (SPI < ZERO) mais ne s'éloignant pas de cette moyenne de plus de deux écarts types de la moyenne (SPI au moins égal à m - 2S').

LIMITE SUPERIEURE DU ONZE DES NOTES.

On désignera par cette note un candidat évidemment supérieur à la moyenne mais ne s'écartant pas de cette valeur d'une façon significative. La limite supérieure s'impose donc et doit être ainsi définie:

ONZE désigne un sujet dont le SPI ne dépasse pas m + 2S'.

FRONTIERE ENTRE DIX ET ONZE DES NOTES

Ces deux notes désignent des candidats supérieurs à la moyenne mais ne sortant pas de l'intervalle de confiance qui vaut deux écarts types (risque 5p.100).

. La limite inférieure de DIX étant m

. La limite supérieure de ONZE étant m + 2S'

. La frontière simple peut être fixée à m + S'.

DEFINITION DES NOTES "MOYENNES".

Elles sont comprises entre m - 2S' et m + 2S'.

. Définition du NEUF : $m - 2S' \leqslant SPI < m$

. " DIX : $m \leqslant SPI < m + S'$

. " ONZE : $m + S' \leqslant SPI < m + 2S'$

Au total :

. Sont définies les notes : 0,9,10,11 et 20: cinq sur vingt et un, il en reste donc seize.

. Sont approchées les définitions des notes 1 et 8,12 et 19.

CRITIQUE DE CES DEFINITIONS.

Il a été pris comme unité servant à définir les 9,10,et 11 des notes, l'écart type de la moyenne. Un statisticien pointilleux pourrait reprocher une faute (serait-ce son seul reproche ?). La faute consiste à établir l'écart type d'une moyenne de population et non d'un échantillon. Celà revient à dire que les candidats qui se sont présentés à l'examen ne représentent pas toute la population et que, par conséquent, la moyenne observée, celle de l'échantillon, fluctue autour de la moyenne "vraie". Le reproche qu'on peut faire à la méthode proposée est de considérer que la totalité des candidats ne représente qu'un échantillon. Que le tout n'est qu'une partie du tout.

Il est vrai que la totalité des candidats n'est considérée que comme un échantillonnage et celà pour la simple raison que tous les candidats, d'une certaine année par exemple, ne représentent qu'une partie des candidats subissant la même épreuve sur un ensemble d'années. En d'autres termes, ce choix implique la possibilité de comparaisons d'une année sur l'autre et, dans la même année, éventuellement, d'un groupe à un autre ou d'un établissement à un autre pour peu que les épreuves soient comparables.

DEFINITION DES AUTRES NOTES

Il reste seize intervalles (notes) à définir.

19, 18, 17, 16, 15, 14, 13, 12
1, 2, 3, 4, 5, 6, 7, 8

Soit huit de chaque côté des "notes moyennes". Pour des raisons de simplicité évidentes, huit définitions restent à établir (de 12 à 19) les huit autres étant exactement symétriques seront définies de la même manière au signe près.

UNE FRONTIERE MAL DEFINIE.

C'est la frontière qui sépare 12 de 11 (ou 8 de 9). En effet il faut à tout prix éviter que la limite inférieure de 12 soit inférieure à la limite supérieure de 11. Si cette condition n'est pas respectée,un intervalle ambigü sera créé dans lequel on pourrait aussi bien attribuer 11 que 12 (ou 8 que 9) ce qui ferait perdre toute signification à ces degrés.

RAPPORT S SUR S'.

La connaissance de la marge de fluctuation de ce rapport, permet d'éviter l'inconvénient signalé.

Soit une épreuve comportant un nombre d'item très proche du minimum admissible: Qe = 36

On sait que $S' = \frac{\sqrt{Po}}{\sqrt{Qe}}$ et, ici, $S' = \frac{\sqrt{Po}}{\sqrt{36}} = \frac{\sqrt{Po}}{6} = \frac{S}{6}$

puisque $\sqrt{Po} = S$.

Puisque $S' = \frac{S}{6}$, on en déduit: 2S' = 0.33S.

La valeur choisie de Qe étant, pratiquement, à la limite inférieure du possible,tout Qe réel sera égal ou supérieur et, par conséquent, l'intervalle de confiance de la moyenne (2S') vaudra moins de 0.33S.

Or m + 2S' définit la limite supérieure du ONZE des notes. Par conséquent, la limite supérieure du DOUZE doit se trouver au moins à m + 0.33S. Dans une épreuve de 36 item, la limite supérieure de onze (qui est par définition la limite inférieure du DOUZE) se confondrait avec la limite supérieure de douze ce qui reviendrait à faire disparaître la possibilité d'attribuer l'une des deux notes.

Au total: la limite inférieure du DOUZE doit se trouver au moins à m + 0.33S.

UN SYSTEME SEDUISANT.

Considérant, d'une part, qu'il y a huit intervalles à couvrir, d'autre part que, de 10 à 19, l'intervalle représenté est de 2S, on peut être tenté de prendre comme unité d'intervalle des "dixièmes de 2S".

L'expression peut sembler manquer de simplicité mais elle propose un moyen de mnémotechnique très simple et, par conséquent, simplificateur(1). En effet, après calcul de 2S qu'on en prenne les dixièmes et :

. dix dixièmes représentent la limite inférieure de 20
. neuf " " " " " 19
. huit " " " " " 18
. sept " " " " " 17
. six " " " " " 16
. cinq " " " " " 15
. quatre " " " " " 14
. trois " " " " " 13
. deux " " " " " 12

Ainsi la limite inférieure de 12 est donnée par 0.2 x 2S = 0.4S ce qui est parfaitement compatible avec la définition de la limite supérieure de 11 qui est située, au plus, à m + 0.33S.

UNE ZONE DIFFICILE A DEFINIR.

C'est l'intervalle qui sépare la limite supérieure de ONZE de la limite inférieure de DOUZE. Dans le cas de 36 item,cette différence n'est pas grande (0.07S) mais dans le cas de 100 item, elle est sensiblement plus importante (0.2S). On a le choix entre trois solutions:

1°- marquer cet intervalle d'un signe spécial (11.5 ou 11^{+}). Le demi point a l'inconvénient de compliquer les calculs de moyennes. Le signe (+) a l'avantage de ne pas rentrer dans des opérations mais ne sert que si on se refuse absolument à choisir entre 11 et 12 ce qui pourrait se concevoir si on envisage des limites de compensation entre épreuves.

2°- élever la limite supérieure de 11 jusqu'à m + 0.4S. Ce qui présente l'avantage de la simplicité et l'inconvénient d'élargir artificiellement la zone des " notes moyennes".

3°- abaisser la limite inférieure de 12 jusqu'à m + 2S'. Ceci présente le double avantage de respecter la signification des notes moyennes et de rendre compte de cette réalité, à savoir que plus est grand le nombre des item et plus étroit est l'intervalle de fluctuation de la moyenne.

Quelle que soit la solution retenue, il n'est pas possible d'en changer si on souhaite comparer des notes et la comparaison est, finalement, leur seule justification.

---

(1) Ou "dixièmes de distribution prévue" à condition de se rappeler que cette distribution vaut ± 2S.

DEFINITION DES 21 DEGRES OU DE 20 INTERVALLES

VINGT----------				SPI	>	m	+	2S
DIX-NEUF--------	m + 2S x 0.9	≤	SPI	≤	m	+	2S	
DIX-HUIT--------	m + 2S x 0.8	≤	SPI	<	m	+	2S x 0.9	
DIX-SEPT--------	m + 2S x 0.7	≤	SPI	<	m	+	2S x 0.8	
SEIZE----------	m + 2S x 0.6	≤	SPI	<	m	+	2S x 0.7	
QUINZE---------	m + 2S x 0.5	≤	SPI	<	m	+	2S x 0.6	
QUATORZE--------	m + 2S x 0.4	≤	SPI	<	m	+	2S x 0.5	
TREIZE----------	m + 2S x 0.3	≤	SPI	<	m	+	2S x 0.4	
DOUZE-----------	m + 2S'	≤	SPI	<	m	+	2S x 0.3	
ONZE------------	m + S'	≤	SPI	<	m	+	2S'	
DIX-------------	m	≤	SPI	<	m	+	S'	
NEUF------------	m - 2S'	≤	SPI	<	m			
HUIT------------	m - 2S x 0.3	≤	SPI	<	m	-	2S'	
SEPT------------	m - 2S x 0.4	≤	SPI	<	m	-	2S x 0.3	
SIX-------------	m - 2S x 0.5	≤	SPI	<	m	-	2S x 0.4	
CINQ------------	m - 2S x 0.6	≤	SPI	<	m	-	2S x 0.5	
QUATRE----------	m - 2S x 0.7	≤	SPI	<	m	-	2S x 0.6	
TROIS-----------	m - 2S x 0.8	≤	SPI	<	m	-	2S x 0.7	
DEUX------------	m - 2S x 0.9	≤	SPI	<	m	-	2S x 0.8	
UN--------------	m - 2S	≤	SPI	<	m	-	2S x 0.9	
ZERO------------			SPI	<	m	-	2S	

On peut distinguer deux manières de calculer les notes:

1°- En appliquant le tableau précédent stricto sensu

. $SPI = \Sigma R_c - \Sigma R^2$

. $SPI = \Sigma R_m - \Sigma R^2 = 0 \quad (\Sigma R_m = \Sigma R^2)$

avec "c" pour "le candidat" et "m" pour "minimal".

. Ainsi, il n'est pas que de comparer SPI à 0±2S (ou 2S')

2°- En évitant l'opération $\Sigma R_c - \Sigma R^2$

. ΣR_c représente la somme des R du candidat

. ΣR_m représente la somme des R minimale pour être reçu.

. Il s'agit alors de comparer ΣR_c à $\Sigma R_m \pm 2S$ (ou 2S'

Au total, les critères qui président à l'établissement de la note sont des résumés objectifs des résultats: moyenne et dispersion. Ce qui est arbitraire c'est la définition de chaque échelon, tout comme l'est le choix même de l'échelle. Mais il semble bien qu'un choix est moins arbitraire que l'autre.

Quelle soit la facilité de l'épreuve, toutes les notes étant définies représentent quelque chose et, l'unité choisie étant l'écart-typ dans toutes elles représentent les mêmes choses.

REPARTITION D'UNE POPULATION NORMALE

Le système de notation proposé est fondé sur l'hypothèse d'une répartition gaussienne. Autrement dit, il est possible de prévoir quelle fraction de la population aura telle note. La répartition observée risque d'être sensiblement différente de ce qui avait été prévu et cette différence sera significative des caractéristiques non de la notation mais de la population. L'histogramme théorique n'est pas un modèle mais une manière de représenter ce à quoi ressemblerait une population répartie au hasard autour de la moyenne. L'histogramme observé permet de montrer objectivement si la répartition est aléatoire ou non et,dans ce cas, qui manifesterait une organisation, dans quel sens se fait l'organisation.

Les notes traduisant des fractions d'écart-type, l'effectif "gaussien" de chacun des degrés peut être ainsi prévu à plus ou moins 0.25p.100 près.

NOTES	EFFECTIFS (environ)	
ZERO	2.5 p.100	2.5 VINGT
UN	1.5 " "	1.5 DIX-NEUF
DEUX	2 " "	2 DIX-HUIT
TROIS	2.5 " "	2.5 DIX-SEPT
QUATRE	3.5 " "	3.5 SEIZE
CINQ	4.5 " "	4.5 QUINZE
SIX	5 " "	5 QUATORZE
SEPT	6.5 " "	6.5 TREIZE
HUIT	7 " "	7 DOUZE

NOTES MOYENNES

NOTES	EFFECTIFS (environ)
NEUF	14.5 p.100
DIX	8 " "
ONZE	7.5 " "

(1 point = 0.1 x 2S)

A- Deux faits peuvent être constatés:(cf.p.256)

1°- le mode prévu est 9 et non 10. Ceci s'explique par le fait que 9 représente un écart de la moyenne de - 2S' alors que le même écart positif se répartit entre 10 et 11. Le seul fait de constater que 9 sur 20 n'est pas la note la plus fréquente a une signification et atteste que la population n'est pas répartie au hasard.

2°- L'histogramme se relève à ses deux extrémités et ceci traduit le fait qu'il a été artificiellement limité à ± 2 écarts-types

B- Cinq histogrammes peuvent être construits aux fins d'analyse: (cf.p.256

1°- Un histogramme sans nuances: il montre que les probalités de n'être pas reçu (49.5 p.100) sont égales à celles d'être reçu (50.5 p.100) àtrès peu de chose près.

2°- On peut établir un histogramme à trois classes: les non-reçus (35 p.100), les reçus (35p.100) et les "moyens" (30 p.100 répartis entre 9, 10 et 11 sur 20). L'observation d'un histogramme à petite classe moyenne serait la preuve que l'examen était ségrégatif: il aurait nettement séparé les reçus des non-reçus avec un minimum de "douteux".

3°- Un histogramme à cinq classes permettrait de séparer arbitrairement les mauvais (5 à 8 : 23 p.10) des pires (0 à 4 : 12 p.100) d'une part et les bons (12 à 15 : 23 p.100) des meilleurs (16 à 20 : 12 p.100) d'autre part; les "moyens" restant évidemment à 30 p.100. La comparaison de l'histogramme réel avec cet histogramme théorique peut aider à se faire une idée plus nuancée de la population contrôlée

4°- Il peut être utile d'aller plus loin en considérant l'histogramme à sept classes où les effectifs théoriques seraient:

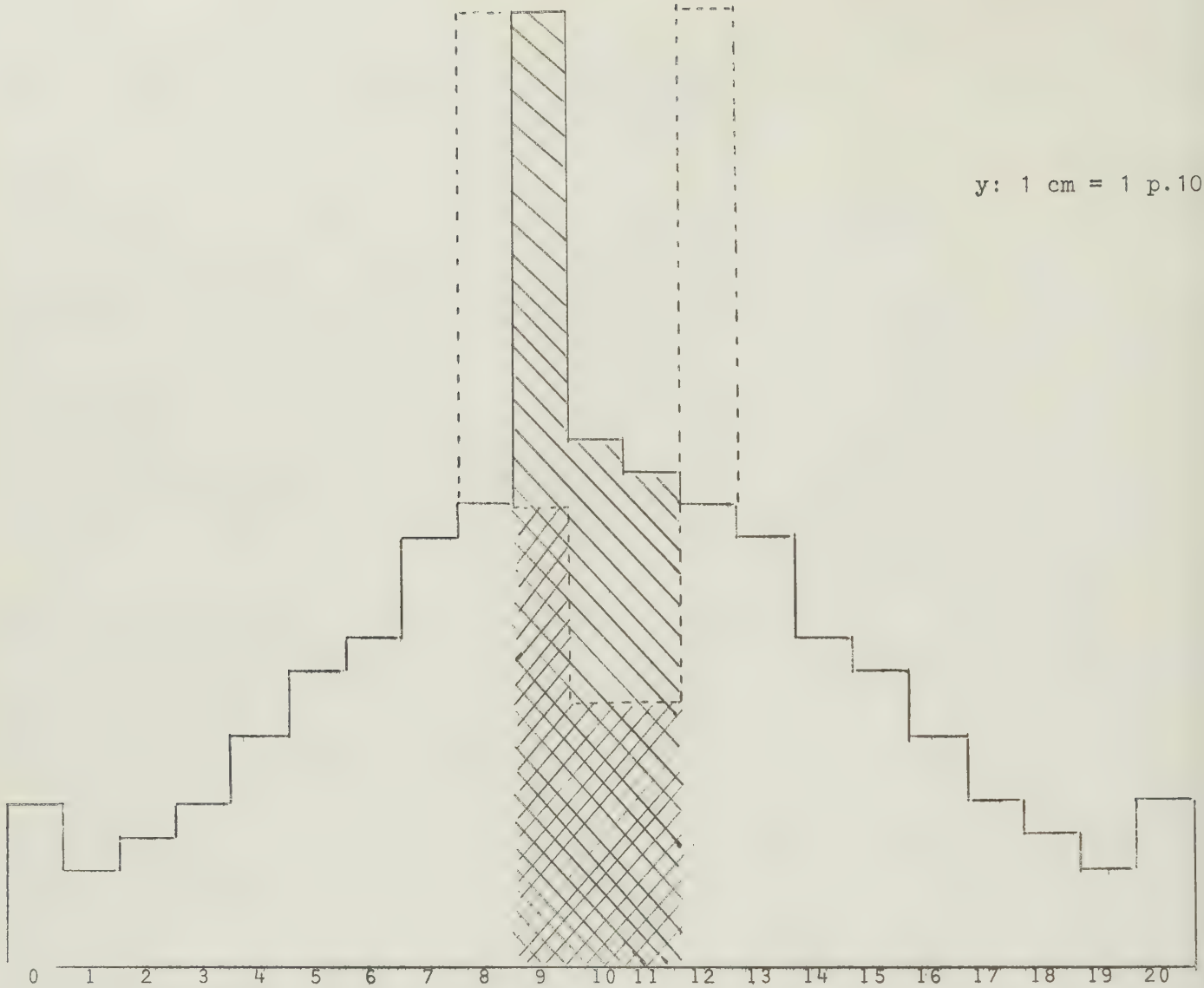
- 30 p.100 de "moyens".
- 23 p.100 d'assez bons (12 à 15) et de médiocres (5 à 8).
- 9.5 p.100 de bons (16 à 19) et de mauvais (1 à
- 2.5 p.100 d'exceptionnellement bons (20) et d'exceptionnellement mauvais (0).

Ce type plus nuancé permettrait de suivre avec précision les progrés d'une population par le "contrôle continu". Une épreuve précédant l'enseignement et d'un niveau "terminal" fournirait un très fort contingent de basses notes tandis que la même épreuve survenant après l'enseignement devrait fournir un fort contingent de hautes notes, un transfert vers la droite s'opérant en cours de contrôle.

5°- Enfin, et pour éviter tout arbitraire telle que la séparation des "bons" et des "très bons" entre 15 et 16, on peut envisager une autre répartition en cinq classes:

- les exceptionnellement mauvais: et les exceptionnellement bons séparés de la moyenne par plus de deux écarts types (2.5 p.100 pour chaque groupe).
- les "moyens" : dans l'intervalle m ± 2S': 30 p.100.
- les bons ou les mauvais en dehors de l'intervalle de confiance de la moyenne et en dedans de m ± 2S : 32 p.100 pour chaque groupe.

y: 1 cm = 1 p.100

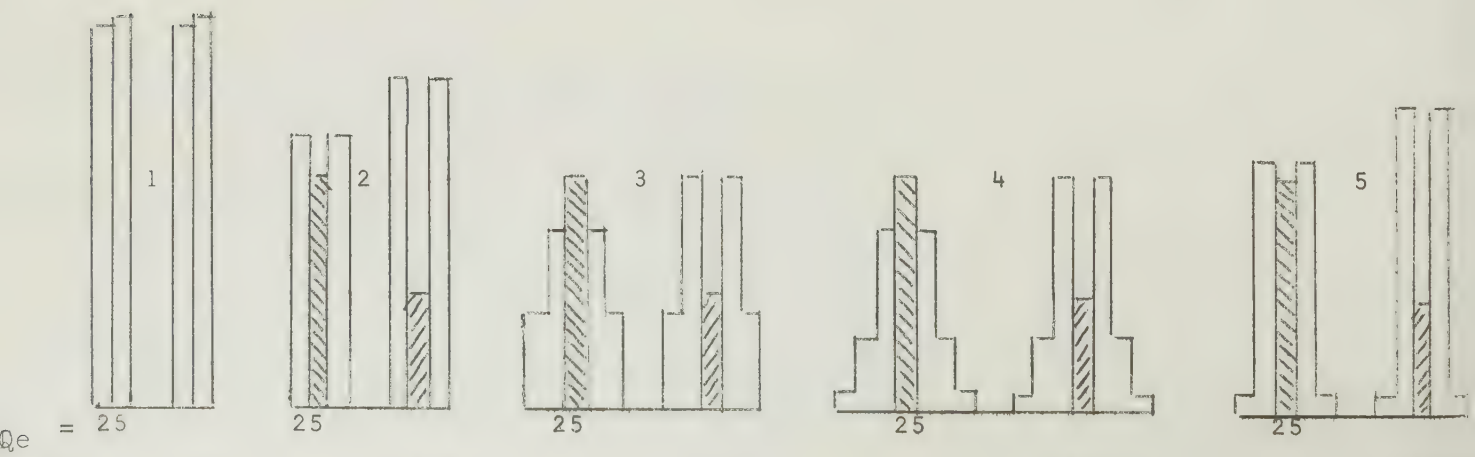


Limité en pointillés: l'histogramme "Gaussien" pour Qe=100

En hachures "notes moyennes" pour Qe=25

En hachures "notes moyennes" pour Qe=100

y: 1 mm = 1 p.100



C - L'histogramme de référence représente un cas particulier.

En admettant que la répartition soit réellement gaussienne, un tel histogramme n'est vraisemblable que dans la mesure où l'écart-type de la moyenne présente une valeur définie: le cinquième de l'écart-type:

$$S' = 0.2 \times S$$

Donc $S' = \frac{S}{5}$ avec $S' = \frac{S}{\sqrt{Qe}}$ et, puisque $S' = \frac{S}{\sqrt{25}}$, Qe = 25 item.

Or, les calculs statistiques sur lesquels est fondée cette méthode, n'ont de sens que pour un nombre d'item au moins égal à 30. Donc l'effectif prévu des notes moyennes, dans une hypothèse gaussienne, sera toujours inférieur à ce qui avait été indiqué.

Pour Qe = 100, on aurait S' = 0.1S. Dans l'hypothèse simplificatrice d'une répartition par classes (histogramme) la moitié des neuf du cas précédent (Qe = 25) deviendraient huit, la totalité des onze deviendraient douze et l'effectif initialement attribué au dix se répartirait, à peu près également, entre dix et onze. Le nouvel effectif des notes moyennes deviendrait:

- Neuf: 14.5 p.100/2 : environ 7 p.100
- Dix + onze: (8 + 7.5) p.100/2 : environ 8 p.100
- L'effectif du huit gagnerait ce que le neuf a perdu et deviendrait 14.5 p.100.
- L'effectif du douze gagnerait ce que le onze a perdu et deviendrait 14.5 p.100.

Au total, l'histogramme de référence deviendrait:

Sept---------- 6.5 p.100
Huit ---------- 14.5 p.100
Neuf ---------- 7 p.100
Dix ----------4 p.100
Onze ---------- 4 p.100
Douze ----------14.5 p.100
Treize ----------6.5 p.100

Un tel histogramme présenterait l'avantage de faire passer l'effectif des moyens (incertains) de 30 p.100 à 15 p.100 et d'amorcer une répartition bimodale avec, d'un côté les reçus certains, de l'autre les certainement non reçus (risque 5 p.100) et un nombre réduit d'intermédiaires et, notamment, de "presque reçus": les neufs sur vingt. C'est à l'examinateur de composer ce standard de référence en proposant un grand nombre d'item.

Au total, cette méthode de notation permet de comparer une réalité objectivement définie à un type standard tout aussi rigoureusement défini: la répartition gaussienne qui n'est plus un modèle mais une référence. Ainsi la note semble bien avoir une signification et deux notes deviennent comparables, leur éventuelle différence pouvant être interprétée.

COMPENSATIONS

Dans le cas d'une notation fondée sur le coefficient R, la séquence des opérations est la suivante:

1°- L'examinateur pose (rédige) les questions et les réduit en item; qu'il choisisse de les présenter sous forme d'item distincts ou sous forme d'un ensemble d'item (questions rédactionnelles) ne change rien de fondamental. Il prévoit les réponses à retenir comme acceptables.

2°- Les candidats répondent

3°- Le noteur accepte ou non chacun des item: établissement du score.

4°- L'ensemble des calculs à faire peut être aisément automatisé: du score à la note.

5°- Chacune des copies est finalement notée.

Il est à remarquer que l'examinateur n'intervient qu'au commencement de la séquence. N'importe qui, pour peu qu'il sache lire et compter peut assurer la suite des opérations. L'intervention de l'examinateur est en revanche indispensable dans le déclenchement du processus: savoir poser des questions (et pourquoi elles sont posées et ce que signifient les réponses faites) constitue son vrai rôle. Le fait qu'il note ne serait pas une déchéance mais ne pas savoir accomplir l'activité qui lui est propre serait une démission.

6°- Le dernier temps appartient au jury. Il lui a été reproché de prendre des décisions contradictoires mais ce fait avait été expliqué par l'insuffisance des informations dont il disposait. La communication du score de chacune des disciplines dont il aura à juger constituerait une information capitale puisqu'il pourrait tout savoir des épreuves: les questions posées, les réponses attendues, la réussite de chacun des item, et le score (en nombre) le SPI et la note de chacun des candidats ainsi que l'histogramme des notes de chaque discipline. L'utilisation de ces informations suppose que les jurés les aient en mains suffisament tôt car ce genre de documents nécessite un certain temps d'analyse pour être utilisable. La gravité des fonctions du jury est telle qu'aucun de ceux qui ont la charge d'y siéger ne peut envisager comme une perte de temps de s'informer avant de juger.

L'avantage supplémentaire du protocole qui consiste à faire aboutir toutes les données "primaires" au jury, c'est qu'il est sûr que les cartes ne sont pas truquées: les notes sont exactement déterminées: aucun jugement ne les a modifiées car c'est au jury de juger.

Dans cette méthode de notation, il est tout à fait possible qu'un candidat soit désigné comme "insuffisant" avec 9 sur 20, parce que son SPI était de deux centièmes de points inférieur au Score Pondéré minimal: il est normal qu'il n'ait pas 10 sur 20. "Arrondir" la note ou recevoir le candidat malgré cette insuffisance serait l'aboutissement d'un jugement, arbitraire de plus, et cet acte est de la compétence exclusive du jury. Les documents fondamentaux ne doivent avoir subi aucune retouche.

On peut penser que le jury aurait avantage à se prononcer très tôt (par exemple au début de l'année) sur le genre de documents qu'il souhaite posséder et sur la présentation qu'il juge convenable. Toujours dans l'hypothèse de l'application de la méthode "RER" jusqu'à l'établissement de la note, on peut supposer que le jury ne délibérera pas sur le sort des candidats qui ont été jugés satisfaisants dans chacune des disciplines: dans le cas d'une présentation binaire des résultats, le tableau des candidats ayant été satisfaisants dans chaque discipline peut être simplement sanctionné, en bloc, par la mention "Reçu".

Tous les autres cas doivent être examinés individuellement et semble-t-il, sur trois sortes de critères: le nombre d'épreuves réussies, le total des notes qui donne une idée de la profondeur de l'échec et le détail des notes qui, seul, peut montrer où a été subi l'échec: tout ceci supposant que les notes soient significatives.

Le jury a parfaitement le droit de critiquer l'épreuve elle même mais il serait mieux que cette critique soit utile donc qu'elle s'exerce sur un projet d'épreuve et non sur les conséquences d'une épreuve mal conçue. Les seules choses que le jury n'ait pas le droit de toucher ce sont les informations qui lui sont fournies: le score et ce qui en découle. Pour tout le reste, il doit être parfaitement libre: il prendra réellement, alors, les décisions arbitraires qu'il a pour fonction de prendre. Schématiquement, les attitudes qu'il peut adopter peuvent se ramener à cinq types fondamentaux:

a) connaissance prise des documents, il déclare faire entière confiance aux notes et ne recevoir que les candidats ayant satisfait à chacune des épreuves: c'est une délibération rigoureuse extrême.

b) il admet de passer outre aux notes de certains candidats sur lesquels il estime avoir des renseignements favorables plus fiables: c'est une délibération individuelle personnalisée.

c) il admet que, seul, le total soit pris en considération: il déclare ainsi que l'échec dans certaines disciplines, ou dans n'importe laquelle, est sans importance pourvu que la réussite dans telle autre, ou dans n'importe laquelle, soit éclatante. Tout caprice ayant été éliminé de la notation, on peut, en effet, accorder une certaine confiance au candidat qui aurait obtenu un 20 sur 20. Rien n'empêche le jury de déclarer systématiquement reçu tout candidat ayant obtenu au moins un 20 sur 20. Inversement, il peut s'opposer à la réception d'un candidat ayant eu un zéro avec la signification précise qui s'attache à cette note:de telles décisions peuvent être prises longtemps à l'avance et rendues publiques ou non. C'est une délibération fondée sur la réussite exceptionnelle.

d) il admet, sans aucune limite, la compensation intégrale en ne prenant que le total en considération: c'est une délibération à laxisme maximal.

e) Il admet la compensation dans certaines limites: c'est une délibération à rigueur tempérée. Informé qu'il est de la signification des notes, il peut ainsi décider devant les notes obtenues à un examen composé de deux épreuves:

. 11 + 9.

- Candidat reçu car ne s'éloignant pas significativement de la moyenne ni dans un sens ni dans l'autre.

- Candidat refusé car doublement incertain

. 12 + 9.

- Candidat reçu car l'échec n'est pas significatif.

. 12 + 8.

- Candidat refusé car l'échec est significatif.

Aux cinq types précédents qui avaient le caractère commun d'être globaux, on peut en ajouter un sixième qui serait ponctuel. Dans un examen à trois épreuves le jury peut prendre les décisions suivantes:

. NOTES/ 12 + 9 + 8

- Candidat reçu à la première épreuve

- Candidat reçu à la deuxième épreuve car l'échec n'est pas significatif.

- Candidat refusé à la troisième épreuve car l'échec est significatif.

Dans un examen à cinq épreuves ce type de délibérations ponctuelles peut beaucoup changer le pronostic. Exemple :

12 + 9 + 8 + 7 + 6

- Candidat refusé à 4 épreuves. Pronostic de la deuxième session: 12 p.100 de chances d'après les statistiques de 1970 et 1971.

- Candidat accepté à la seconde (échec non significatif) les chances d'être reçu sont alors de 47p.10 donc, en gros, multipliéss par 4.

Dans tous les cas, ce qui est important c'est que 9 sur 20 reste 9 sur 20 au moins jusqu'à l'intervention du jury.D'ailleurs,les notes n'ont pas à être changées par le jury: s'il accepte le système de notation proposé il ne peut logiquement modifier une note que dans la mesure où une erreur est décelée. Son rôle, qui n'est en rien minimisé, n'est pas de fournir une tâche de comptable mais de faire en sorte que les notes soient significatives et de décider éventuellement de recevoir des candidats tels qu'ils sont: avec des notes insuffisantes. Il a parfaitement le droit d'accepter un candidat 9 + 9 + 9 + 9 + 9 en jugeant que le quintuple échec n'est pas significatif, sans pour autant déclarer que 9 x 5 = 50.

Au total, c'est au jury qu'il doit revenir de décider des "compensations", libre à lui de se prononcer tôt mais il semble que, toute note égale ou inférieure à huit signifiant une insuffisance, les décision arbitraires de réception ne peuvent être que des critiques du résultat de l'examen fondées sur des renseignements favorables plus fiables, par exemple, une fiche personnelle du candidat apportant des informations et non des impressions.

NOTATION ET COEFFICIENT NORMATIF

Tout ce qui précède, dans ce chapitre, était fondé sur le coefficient de réussite. Toutes les règles peuvent être appliquées à l'utilisation du coefficient normatif à la seule condition que ce coefficient soit l'expression d'une exigence rationnellement établie sans quoi tous les calculs seraient vains.

Il est une circonstance où le coefficient de réussite semble d'application dangereuse: la deuxième session (septembre traditionnellement). En effet, les candidats sont, dans l'ensemble, moins bons et la prise en considération du coefficient de réussite risque fort d'abaisser le niveau de l'épreuve par rapport à l'ensemble de la population considérée, puisque les meilleurs en sont exclus. Deux solutions s'offrent:

a) rendre la seconde session obligatoire pour tout le monde sans remettre en question la réussite obtenue en première session. Ceci aurait l'avantage de situer les vrais candidats par rapport au niveau réel de la population et éviterait l'inconvénient de rendre artificiellement difficiles des item qui seraient manqués par un échantillon sélectionné précisément sur son manque de réussite. Il n'est pas certain que cette solution soit acceptée. Il en reste une autre:

b) appliquer à la deuxième session les coefficients observés à la première. Ceci nécessite quelques précautions, notamment dans le domaine de la discrétion. Si les item de la première session ont été suffisamment nombreux et s'ils sont restés absolument secrets (et ce n'est pas impossible) les candidats ne pourront pas se les rappeler. Il en serait autrement sur une question rédigée après publication de la bonne réponse type. Il est encore possible d'utiliser les coefficients observés dans les examens partiels. Enfin, l'analyse détaillée des résultats d'un examen semble montrer que certains item, un assez petit nombre, auraient fournis à eux seuls des résultats comparables. Il est donc possible d'utiliser ces item caractéristiques de la première session dans la seconde en conservant leurs coefficients qui seraient devenus normatifs et rien n'empêche d'ajouter des item nouveaux évalués sur leurs coefficients de réussite.

CONCLUSION

Il semble que la méthode proposée atteigne ses objectifs: donner un sens aux résultats et une signification aux notes.

Un histogramme de répartition des notes n'aurait pas le sens qu'on est souvent tenté d'y voir: une manifestation caractérielle du "correcteur". L'obligation faite à l'examinateur de proposer les réponses en même temps que les questions oblige à donner un sens défini à l'épreuve qui risque alors d'être critiquable. Il est préférable de composer une épreuve critiquable que de fournir un magmas tellement inorganisé qu'il ne serait même pas digne de critique et la possibilité de critique est un facteur de progression.

L'option fondamentale à prendre pour un examinateur en ce qui concerne l'établissement du résultat se situe au niveau du choix: les Résultats seront-ils Etablis sur le coefficient de Réussite (Méthode R.E.R.) ou non ? Répondre "non" implique la détermination normative de la valeur de chacun des item et suppose qu'on ait des raisons de le faire et, par exemple, d'en être capable. Au delà, l'examinateur a cessé son rôle jusqu'au jour de la réunion du jury et, sur ce point, on peut discuter entre deux types de composition de jury selon qu'il serait formé d'examinateurs ou qu'il s'assurerait de leur participation dans un groupe plus étendu. Dans le cas, qui semblerait heureux, où le service d'examens serait relativement indépendant du service d'enseignement, une confrontation, au sein du jury, entre Professeurs et Examinateurs pourrait être fructueuse par les points de vue assez différents qui seraient manifestés. L'un des avantages majeurs de cette solution, serait l'indépendance des juges vis à vis de la procédure d'instruction.

Entre la composition de l'épreuve et la réunion du jury, que les opérations soient humaines ou automatiques, elles obéissent à des règles strictes et aboutissent, dans chaque épreuve:

a) à un résultat individuel: reçu ou non.
b) à une note très fidèlement représentative de la situation d'un candidat dans la population examinée.

Le caractère rigide des opérations est une nécessité si on estime souhaitable que l'"humain" (le jury) soit convenablement informé aux fins de juger aussi justement que possible. Le jury est largement réhabilité.

C - DEFINIR LES OBJECTIFS

" Comment faire ?
Quoi faire ? "

Il ne suffit pas,pour qu'un examen soit bon, que ses résultats soient établis avec sensibilité, fidélité et sans dérive. Il faut encore qu'il soit juste. L'une des manières de le rendre juste, et des plus efficaces, consiste à définir la nature de ce qu'il doit mesurer et cette description peut être précise.

Les études, quelles qu'elles soient, mobilisent suffisamment de temps et de moyens pour qu'il ne soit pas inutile de savoir à quoi elles mènent. Lorsque le but en aura été défini, et à ce moment seulement, il sera possible dechoisir les moyens d'y parvenir. Les moyens étant nombreux, leur choix n'est pas toujours aisé; la possession d'un instrument de contrôle efficace permettra de juger de leurs valeurs respectives à condition expresse que soit défini ce vers quoi ils sont censés mener: un but précis.

CHAPITRE V : UTILITE D'UN EXAMEN

CHAPITRE VI: OBJECTIFS D'UN EXAMEN

CHAPITRE VII:EXAMEN ET CONTROLE DES CONNAISSANCES

CHAPITRE VIII : OBJECTIFS DEFINIS D'UN EXAMEN

CHAPITRE V

UTILITE D'UN EXAMEN

C - DEFINIR LES OBJECTIFS

Le seul fait qu'il s'agisse d'un acte imposé par l'usage et par les réglements rend l'examen suspect. Pour les candidats, c'est une formalité pénible dont le sens est discutable et son existence même, pour les enseignants, c'est une corvée (acte gratuit que doit le vassal à son suzerain). La question s'est déjà posée, brutale, pourquoi ne pas supprimer les examens? C'est sûrement ce qu'on aurait de mieux à faire s'ils ne servent à rien. Dans cette hypothèse, il faudrait oeuvrer pour que change la loi. Si, au contraire, une utilité leur est reconnue, il faut le définir, rejeter tout examen inutile et faire en sorte que l'acte soit en conformité avec son but.

AUTORITE ET EXAMEN 266
POINT DE VUE DU CANDIDAT 267
POINT DE VUE DE L'ETUDIANT 268
POINT DE VUE DE L'ENSEIGNANT 269
COMPETITION 270

AUTORITE ET EXAMEN

En règle, l'autorité c'est directement ou indirectement, l'administration : les Pouvoirs Publics. De même qu'on attend d'elle la définition des poids et mesures ou de la monnaie et diverses sortes de garanties de sécurité, on peut en espérer que les diplômes dont elle a le monopole de l'attribution aient une signification : que l'impétrant ait une capacité correspondante à la réputation qu'on lui attribue.

Si l'autorité était privée, l'inadéquation du signe à la réalité aurait, certes, quelques inconvénients, mais finirait par disparaître consécutivement à la mauvaise réputation qui s'attacherait, à la longue, à un produit de mauvaise qualité. L'inconvénient possible de l'acte d'autorité publique, en la matière, c'est qu'une éventuelle déviation de la démarche par rapport au but choisi, risque d'aboutir, non à l'échec mais à la dénaturation de l'objectif : qu'il y ait substitution de la forme au fond. Que l'apparence tienne lieu de réalité. Un examen, ainsi privé de fond, jouerait peut-être le rôle d'une cérémonie initiatique mais ne serait absolument pas un contrôle dans le sens français du terme.

Au total, l'exigence d'un examen, exprimée en terme de droit , ne doit pas être envisagée sous le seul aspect de la matérialité d'un fait, mais encore, mais surtout, sous celui d'une signification.
A quelles réalités doit correspondre le signe ?

POINT DE VUE DU CANDIDAT

Pour un candidat, la définition d'un examen est simple, très claire, c'est le but à atteindre. Il aspire à un grade : que ce terme représente un échelon, une situation(sens statique) ou le franchissement d'un pas. Qu'il s'agisse d'un diplôme sanctionnant un curriculum scolaire ou une autorisation de progression, en quelque sorte hiérarchique, d'un niveau d'étude à un autre jugé supérieur.

La seule réponse à laquelle tienne un candidat est : "oui". La seule qu'il craigne est "non". La sanction(décision irrévocable) n'est que la constatation d'un état de fait : accepté ou refusé.
La seule procédure d'appel qui puisse être éventuellement envisagée est une autre tentative. La question, réellement fondamentale, qui se pose est de savoir si la sanction est juste. Si le refus est l'expression d'une réelle incompétence et l'acceptation la manifestation d'une compétence réelle.
Il est nécessaire, pour répondre à cette question de la justice, de savoir si les questions posées et les réponses fournies, d'une part, sont en rapport, et, d'autre part, si le savoir prouvé par les réponses est lui-même en rapport avec le but des études entreprises. En d'autres termes, il faut savoir si l'épreuve est probante et probante de quoi.

L'examen, vis-à-vis du candidat, doit s'envisager en terme de candidat : a-t-il oui ou non atteint son but ? Personne à ce sujet, n'attend d'impression ou de sentiment,mais une certitude.

POINT DE VUE DE L'ETUDIANT

Ce n'est pas une sanction qu'attend l'étudiant mais une correction. Cette catégorie d'humains, comme d'autres, peut se nourrir d'illusions. Il peut être de bonne foi, s'imaginer savoir et ne pas savoir. S'il a réellement étudié, il sait : mais sait-il assez, et surtout sait-il bien ? Il est en droit d'attendre qu'une erreur de méthode, d'interprétation, de compréhension .... etc, lui soit dénoncée. C'est bien souvent, la confrontation avec une difficulté qui permet son identification. Une difficulté identifiée a de plus grandes chances d'être surmontée qu'une inconnue.

Un étudiant n'est pas désintéressé : un jour, il sera candidat. A ce sujet, son ambition est de faire prendre en compte ce qui peut être porté à son actif et d'effacer ce qui pourrait être inscrit à son passif. Que des critiques lui soient faites, bon. Qu'il en tienne compte, entendu. Mais, à partir de preuves de compétence, fournies, il n'entend pas qu'on puisse lui tenir rigueur, le jour venu; de contre-preuves contradictoires. Ce qui est attendu, c'est une espèce de fonction de filtre permettant d'inscrire une fois pour toutes à l'actif ce qui aurait été fourni, ne serait-ce qu'une fois. Quant au reste, que l'oubli l'efface et ceci pose à l'examinateur une question dont il serait vain de cacher la difficulté.

POINT DE VUE DE L'ENSEIGNANT

Qu'on imagine pour simplifier, que celui qui enseigne n'est pas l'examinateur. Dans de tels cas, qui ne sont pas toujours imaginaires, la sanction appliquée à "ses" candidats touche l'enseignant.
Flatté d'un bon résultat(cest-à-dire d'un pourcentage élevé de reçus), il sera froissé, vexé, voire révolté, devant un fort contingent d'échecs. Une considération plus attentive des résultats l'entraînera, en général, à prendre une attitude plus nuancée. A l'échelon des cas individuels, il ressentira, qu'il le manifeste ou non, satisfaction ou déception. L'enseignant se conduit alors comme les candidats. Peut-on le lui reprocher ?

En faisant abstraction d'une "sympathie" bien naturelle entre gens qui ont travaillé ensemble et dans le même but, sans aller trop chercher jusqu'à quel point les rapports enseignés/enseignants se rapprochent de sentiments filiaux ou parentaux, il n'en demeure pas moins que le bonheur de l'enseignant ayant "placé" tous ses candidats ressort du sentiment du devoir accompli. A gros effectif, bon enseignement et à bon enseignement, bon enseignant. Si les critiques des enseignants vis-à-vis de l'examen ne sont pas nécessairement mieux fondées que celles du candidat, elles ne sont pas, non plus, forcément stupides.

Estimer, lorsqu'on est enseignant, que les sujets d'examen n'étaient pas explicitement au "programme", par exemple, cela pose la question de sa lisibilité. Penser que la "correction" était sans rapport avec les questions posées, cela pose une interrogation quant à leur ambiguïté. Croire que ceux qui sont reçus auraient plutôt dû ne pas l'être et inversement, cela peut faire penser que l'un des deux jugements est faux. Constater que la plupart de ses ouailles sont refusées amène l'enseignant à réfléchir sur la qualité de ce qu'il a fait. Mais n'est-il pas trop tard ? Au-delà de la vexation du pronostiqueur contredit n'y a-t-il pas un réel malaise dont la cause est l'incertitude ? Si des vérifications préalables, de nature semblable, mais réellement semblable, avaient été faites, les pronostics eussent sans doute été plus justes et les futurs candidats en danger, dûment avertis. Elles auraient même permis, qui sait, de déceler un défaut dans l'enseignement au cas où il n'aurait pas été parfait. Pour cela, il faudrait qu'un examen, que chacune des épreuves d'un examen, soit défini en termes tels que chacun puisse les comprendre de la même manière.

Au total, et pour un enseignant, la question de l'examen se pose en termes d'enseignement.

COMPETITION

En France, l'autorité en la matière est représentée par l'Administration qui exige des examens pour délivrer un diplôme qualificatif. Il semble bien difficile de lui donner tort sur ce point. Elle exige que le contrôle-vérification précède le contrôle-estampille et c'est logique. Si on considère l'enseignement comme une entreprise destinée à fournir un produit, le contrôle de la conformité du produit s'impose comme dans toute entreprise. Cet argument en faveur de la survie de l'examen semble décisive et ce n'est pas pour réclamer sa mort qu'il faut en faire le procès, mais pour en obtenir l'amélioration.

Il était une fois un horloger tellement habile qu'on n'avait jamais vu de mémoire d'homme, horloges plus parfaites. Son succès fut tel que, bientôt, l'échope devint atelier et l'atelier, usine. Le souci constant de l'ancien artisan devenu chef d'entreprise était que ses horloges fussent toujours irréprochables. Il créa un rigoureux service de contrôle puis fit faillite. Etonné que son succès aboutît à une catastrophe, il s'enquit auprès de son heureux concurrent de ce qui avait bien pu arriver. "Très simple, répondit-il, vos horloges ne valent rien". "Mais mon service de contrôle ?" "Justement, pour qu'une horloge soit reconnue bonne, il suffisait qu'elle fût jaune". La morale de cette histoire, c'est que la principale qualité d'une horloge n'est pas sa couleur.

Loin de devoir être restreint, le rôle de l'examen devrait être élargi, permettant d'attester que le produit fini possède les qualités qu'on en attend. Du diplôme qualificatif, le moins qu'on puisse attendre, c'est que les qualités qu'il garantit soient définies. Un examen attestant d'une progression ne peut être concevable que vis-à-vis d'un plan de progression défini. Enfin, la qualité d'un produit fini étant, dans une large mesure, dépendante de la manière dont il est élaboré, il convient encore de définir ce qu'on attend de l'enseignement pour savoir si les buts sont atteints. L'examen doit concerner des candidats, des étudiants, et des enseignants, les objectifs doivent être définis : étudiants et enseignants poursuivant le même objet, ils sont en compétition.

CHAPITRE VI

OBJECTIFS D'UN EXAMEN

C - DEFINIR LES OBJECTIFS

Quel qu'il soit, un examen a pour fonction de révéler si le but proposé est atteint ou non. La précision de cette révélation dépend elle-même de la précision avec laquelle le but a été décrit. Chaque épreuve est grossièrement identifiée par la nature de ce qu'elle vérifie et qu'on appelle généralement discipline. Mais, en termes d'examens, il n'est pas suffisant de définir une épreuve par une discipline: "Anatomie" ni même par l'une de ses subdivisions : " Physiologie du coeur ". Pour chacune de ces rubriques, dont l'utilité n'est pas toujours contestable, il faut aller plus loin et déterminer ce dont il faut être capable: connaître, comprendre, savoir appliquer ...etc. Ainsi, peut être dégagé un plan simple et universel, applicable à chaque discipline et permettant de savoir à l'enseignant quoi enseigner, à l'étudiant quoi apprendre et à l'examinateur quoi vérifier. Il est difficile de vérifier quelque chose dont on ne sait rien,qui n'est pas décrit puisqu'on ne peut pas comparer un produit avec un modèle non défini.

LES DISCIPLINES CONSIDEREES COMME OBJECTIFS 272
UTILITE D'UNE NOMENCLATURE D'OBJECTIFS 276
QUALITES REQUISES D'UNE NOMENCLATURE D'OBJECTIFS 277
CLASSIFICATION DES OBJECTIFS (d'après BLOOM) 278
. CONNAISSANCES-MEMOIRE 278
. CAPACITES INTELLECTUELLES ELEMENTAIRES (compréhension application) 279
. OBJECTIFS PLUS AMBITIEUX :Analyse,Synthèse et critique 281

LES DISCIPLINES CONSIDEREES COMME OBJECTIFS

Une discipline, c'est ce qui est enseigné. Dans une perspective d'enseignement, ce terme peut se voir attribuer plusieurs sens : une "matière" et une certaine démarche intellectuelle. Quel que soit le niveau, primaire, secondaire ou supérieur, l'histoire et la physique représentent bien deux "matières" distinctes : leur objet n'est pas le même et deux méthodes différentes : l'histoire n'est pas une science expérimentale. En fait, la discipline est un ensemble de connaissances à acquérir, groupées sous la même rubrique, souvent enseignées par la même personne, pour un groupe donné et dans un temps donné, sanctionnées par une épreuve unique.

Pour peu qu'un stade réellement élémentaire soit dépassé, la définition d'une discipline apparaît plus arbitraire que logique. Ainsi, beaucoup de processus vitaux dont l'étude est le fait de la biologie ou de la physiologie ne peuvent être compris qu'à partir de connaissances qui sont, traditionnellement, le fait d'autres disciplines : physique, chimie, histologie, anatomie, etc... L'arbitraire de telles divisions est un fait mais pas nécessairement une absurdité. Ainsi, dans l'exemple de la physiologie, il est aussi indispensable de connaître certaines autres disciplines que de savoir lire, écrire ou compter : il n'empêche que la physiologie est définie par son objet, tout à fait différent de celui de l'anatomie et dont l'étude nécessite des moyens d'investigation appropriés. L'arbitraire de maintes séparations entre disciplines médicales est évident. Ainsi, la plupart des maladies nerveuses, la rhumatologie et la traumatologie en presque totalité pourraient fort bien se concevoir comme une entité définie, intellectuellement satisfaisante : la pathologie de l'appareil loco-moteur. Une telle définition ne serait pas nécessairement intelligente. Des différences de méthode d'investigation et de traitement justifient la répartition de ces différentes sortes de malades dans des services différents et même, éventuellement, leur circulation d'un service à l'autre en fonction du stade d'évolution de la maladie. L'arbitraire est justifiable sur des données concrètes et, dans la mesure où les études de médecine ont un objet concret, une justification concrète semble préférable à une justification abstraite même intellectuellement, plus logique.

Dans n'importe quel ordre d'étude, il ne semble pas possible d'échapper à ce genre de contrainte. Cependant, l'arbitraire n'exclut pas une certaine logique et, dans la mesure où une certaine étude est justifiée dans plusieurs disciplines, il serait préférable de ne pas faire cinq fois la même au détriment de telles autres qui ne seront jamais faites.... faute de temps. Répartir avec précision les objectifs d'enseignement, sous forme de "matière" définie, cela constitue l'un des actes fondamentaux de l'élaboration d'un programme pris dans son sens de "curriculum". Finalement, un enseignant donné est chargé de l'exécution d'une tâche d'enseignement définie par un catalogue, dont le contenu peut être, dans une certaine mesure, arbitrairement défini. Peu importe que la rougeole soit considérée comme élément de pédiatrie ou de pathologie infectieuse, que les paralysies traumatiques soient enseignées en traumatologie ou en neurologie : ce qui importe, c'est que ces éléments soient enseignés. D'une année à l'autre, le catalogue des objectifs confiés à un enseignant peut changer dans la nomenclature de ses rubriques et cela n'a pas beaucoup d'importance. Ce qui est important, c'est que si cent objectifs ont été choisis, il n'y en ait pas cinquante seulement d'atteints parce que chacun d'eux a été visé deux fois et cinquante oubliés. Si c'est la rubrique qui est choisie comme unité d'enseignement, il reste à la définir en termes d'enseignement.

Il est nécessaire de classer les objectifs élémentaires dont l'ensemble constitue l'unité définie en termes de rubrique de catalogue. Il s'agit de définir ce que voudra dire "savoir" vis-à-vis d'un objectif donné, et par conséquent, de déterminer le niveau de l'enseignement et celui de l'épreuve qui devra vérifier que l'objectif est réellement atteint. Pour chaque rubrique il est nécessaire de se poser la question de l'utilité de telle forme de connaissance. A la question, on répond oui ou non.
Par exemple, l'objectif est d'apprendre la table de multiplication.
On peut au moins se poser trois questions :
a) Faut-il savoir la valeur du produit de n'importe quel nombre entier à un chiffre par n'importe quel autre nombre entier à un chiffre ?

b) Faut-il comprendre ce qu'est une multiplication ?

c) Faut-il savoir faire une multiplication entre n'importe quels nombres entiers ?

Trois formes de connaissance sont possibles à propos d'un objectif défini par sa variété(ici la multiplication) : connaissance factuelle, compréhension, application. L'objectif étant "apprendre la table de multiplication" on doit répondre oui à la question a) et non aux deux autres. Ce qui n'implique pas que savoir faire une multiplication est à considérer comme un objectif définitivement inintéressant. Chacun sait que dans cet exemple, il est impossible d'atteindre l'objectif c) sans avoir préalablement atteint le a), que l'objectif a) est implicitement ajouté à l'objectif c) et qu'il n'est pas indispensable de passer par l'intermédiaire b) pour réaliser la séquence (le programme) "a-c" : un court-circuit est possible. Possible ne signifie pas obligatoire.
Soient deux enseignants à qui se trouve proposé l'objectif c). L'un et l'autre seront vraisemblablement du même avis au sujet de la nécessité de a), mais il est possible que l'un veuille atteindre le b) et non l'autre, chacun ayant de bonnes raisons. Qu'adviendrait-il de leurs élèves, subissant la même épreuve, si la connaissance de b) s'y trouve exigée ? La liberté d'interprétation laissée par la rubrique serait seule responsable des échecs expliqués par l'absence de réponses aux item b) de l'épreuve. Pour caricatural que soit un tel exemple, il n'en recouvre pas moins une solide réalité : les enseignants ont plus de chance d'atteindre un but connu qu'un but inconnu. Dans la mesure où les enseignants ont un rôle à jouer dans l'acquisition de connaissances par leurs élèves, il est à craindre que l'impossibilité d'atteindre les objectifs par les uns n'entraîne la même impossibilité pour les autres.

Une rigoureuse définition des objectifs d'enseignement n'est pas sans influence sur la forme même de l'examen. Soit l'exemple de la multiplication : si le <u>seul</u> but de l'épreuve est de vérifier que les candidats savent faire une multiplication entre n'importe quels nombres entiers, l'usage de tables n'est pas interdit par la logique. L'interdiction de l'usage des tables signifie qu'on veut explorer en même temps deux acquisitions : une connaissance factuelle(la table de multiplication) et une capacité d'application. Deux conceptions semblent défendables :

1°) j'autorise l'usage des tables : l'essentiel c'est de "savoir faire" et à la longue, l'utilisation finira par entraîner l'enregistrement. C'est à force d'application que la connaissance factuelle se trouvera "secondairement" atteinte comme je connais beaucoup de valeurs numériques(tels des numéros de téléphone) à force de les utiliser.

2°) J'interdis l'usage des tables : savoir faire une multiplication, c'est pouvoir la résoudre dans n'importe quelle situation sans aucun autre secours que sa mémoire.

La seconde position généralement admise, implique un contrôle portant exclusivement sur la connaissance des tables sans quoi, en présence de résultats faux, on ne pourra jamais savoir s'ils sont la conséquence d'une faute de mémoire ou d'une faute de méthode.

UTILITE D'UNE NOMENCLATURE D'OBJECTIF :

Il était une fois un homme qui aimait tellement les chemins de fer, que, pendant de longues années, il accumula péniblement une somme considérable pour s'offrir le voyage de ses rêves. Le jour où il s'estima pourvu, il s'en fut à la gare la plus proche et demanda son billet.

- "où voulez-vous aller ? lui demanda l'employé.

- aux Antilles.

- Prenez l'avion ou le bateau : on ne peut pas aller aux Antilles en chemin de fer".

La morale de cette histoire, c'est que choisir les moyens avant le but expose à quelques désillusions. Ainsi faut-il répondre d'abord à la question "quel but se propose-t-on d'atteindre ?"
avant d'envisager les moyens qui peuvent y mener et de choisir les mieux appropriés.

QUALITES REQUISES D'UNE NOMENCLATURE D'OBJECTIFS

Deux qualités sont essentielles : l'universalité et la simplicité. A cet égard, la classification de BLOOM semble excellente.

Cette classification a été établie par un comité de 34 enseignants ayant travaillé ensemble pendant cinq ans (1949-1953). Chacun était attaché à une Université différente au sein de laquelle ils poursuivaient le travail de recherche et de critique. Lorsqu'en 1953 l'ouvrage fut estimé avoir une certaine valeur, il fut l'objet d'une diffusion "confidentielle" auprès de 1.000 enseignants auxquels il fut demandé suggestions et critiques. C'est après examen de ces renseignements supplémentaires que fut édité l'ouvrage (1956).

En ajoutant que cette classification fut largement utilisée dans tout le continent américain, et qu'elle l'est toujours, on peut en souligner ainsi le caractère universel d'autant qu'il est certain que cet ouvrage est connu en Belgique, probable qu'il le soit aussi au moins, en Hollande et dans les Pays Scandinaves.... quelques exemplaires ont même été vendus en France.

La simplicité apparaît d'emblée lorsqu'on s'aperçoit que tous les objectifs du domaine cognitif se trouvent finalement ramenés au nombre modeste de 21. Universalité et simplicité ne seraient que des qualités, en quelque sorte, esthétiques, si ne s'ajoutait l'efficacité. Le niveau de généralité dans chaque rubrique est tel que le risque de fragmentation de l'enseignement est minime. En fait, cette méthode de classification est pratiquement inutilisable pour un enseignant qui ne réfléchirait pas : ce n'est pas une liste de recettes mais une suite de thèmes de réflexion permettant l'adaptation à n'importe quelle discipline en rapport avec le domaine cognitif. C'est enfin,une méthode objective puisque les objectifs définis le sont en termes de comportement qui peuvent être observés, décrits et classés.

Toute classification présente un certain degré d'arbitraire. Tout comme une autre, celle-ci présente un certain aspect conventionnel. Il est indispensable pour l'accepter, de se rappeler qu'elle n'est pas fondée sur une théorie psychologique mais sur des comportements accessibles à l'enseignant et acceptables par lui, quelles que soient ses conceptions sur les mécanismes d'apprentissage, au prix éventuellement, d'une minime contrainte l'obligeant à admettre une définition. A ce prix, chaque mot ayant un sens précis, et le même pour tout le monde, il est vraisemblable que les possibilités d'entente seront plus grandes.

TAXINOMIE DES OBJECTIFS

I - CONNAISSANCES

- informations acquises, mémoriser, associer.
- principes du contrôle : pur et simple comportement de remémoration; rappel des connaissances sous une forme sensiblement identique à celle de l'information fournie.
- classement :

I.10 : connaissances d'INFORMATIONS ELEMENTAIRES, ponctuelles, littéralement : atomiques

11 : TERMINOLOGIE : vocabulaire, symboles.

12 : FAITS, connaissances factuelles.

I.20 - Connaissances des MOYENS d'utilisation des informations élémentaires "type : recette".

21 - CONVENTIONS

22 - RAPPORT DANS LE TEMPS entre phénomènes à un instant donné. tendance évolutive d'un phénomène.

23 - CLASSIFICATION, STRUCTURE CONCRETES

24 - CRITERES

25 - TECHNIQUES

I.30 - Connaissances des ensembles conceptuels : ABSTRACTION

31 - PRINCIPES ET LOIS

32 - THEORIES

II - COMPREHENSION

- Saisir(appréhender) le sens d'une communication, mais pas nécessairement toute sa portée.

- Principe du contrôle : demander une réponse différant formellement de l'information.

- Classement :

II.10 - TRANSPOSITION : TRADUCTION

20 - INTERPRETATION : conclure, résumer, distinguer l'essentiel de l'accessoire, le certain de l'incertain

30 - EXTRAPOLATION ou INTERPOLATION : prévision

III - APPLICATION . Résultat du libre choix de l'abstraction appropriée à la résolution d'un problème. L'utilisation correcte qu'une abstraction exigée attesterait de compréhension.

- Principe de contrôle : utiliser une situation nouvelle pour le candidat afin que le choix ne soit pas la répétition littérale d'un schéma déjà connu. L'analogie n'est pas exclue. Le comportement du candidat est au maximum :

a) La motivation du choix de l'abstraction(application)

b) L'indication d'opération(solution)

c) La présentation de la réponse

- réalisation du contrôle

1°) situations :

A - situation fictive

B - situation inconnue du candidat(réellement inconnue) Ce peut être la version simplifiée d'une situation qui ne sera présentée qu'ultérieurement dans le curriculum.

C - situation connue, présentée sous un aspect nouveau, déconcertant(réellement nouveau)

2°) comportements :

On peut choisir dans la séquence a-b-c- précédente, quatre types :

1- a-b-c- : solution développée et motivée

2- b-c- : solution développée

3- c- : réponse seule

4- a : motif seul (exercice d'application pure)

Au total, douze schémas élémentaires peuvent être conçus

- classement : il est fondé sur la classification des disciplines elles-mêmes.

IV. ANALYSE . Décomposition de l'organisation d'une communication ou d'un matériel.

Trois fonctions peuvent lui être attribuées :

1°) c'est une fin en soi

2°) une étape, permettant une meilleure compréhension

3°) une étape indispensable avant la critique.

Principe du contrôle : présentation d'un matériel nouveau, afin d'interdire des commentaires pseudo-analytiques qui ne seraient que des connaissances. L'absence de familiarité ou une défaillance de mémoire peuvent être compensées par la fourniture extemporanée d'informations.

La forme la plus adéquate est la Q.C.M. qui permet de suggérer les erreurs fréquentes et dont la signification du choix peut être établie.

- Classement.

IV. 10 - Recherche des ELEMENTS

IV. 20 - Recherche des RELATIONS ENTRE ELEMENTS : exemple : de cause à effets.

IV. 30 - Recherche des PRINCIPES D'ORGANISATION.

V - SYNTHESE

- Construction d'un tout à partir d'éléments ou de principes connus en vue de déduire des conséquences inconnues.

- Principe du contrôle :

1° - Liberté de pensée et d'action : ni point de vue, ni conclusion imposés.

2° - Nécessité d'un temps suffisant : souvent long.

3° - Nécessité souvent inéluctable de se contenter d'une seule production jugée représentative.

4° - Difficultés très grandes dans l'évaluation : les résultats comptent moins que les moyens d'y parvenir.

- Classement :

V. 10 = COMMUNICATION originale dans un but précis(informer, décrire, émouvoir, persuader... qui ?

V.20 : Elaboration d'un PLAN D'ACTION

V.30 : Constitution d'un ENSEMBLE DE RELATIONS ABSTRAITES.

VI - CRITIQUE

- Application de critère

- Principe du contrôle : expression d'un jugement(d'un jugement et non d'une opinion) consécutivement à 1' application de critères imposés ou non.

- Classement.

VI-10 : CRITIQUES INTERNES : rigueur, cohérence, justesse, exactitude ...

VI.20 : CRITIQUES EXTERNES

a) moyens et fins

Les moyens sont-ils adaptés à leur fin (s) ?

Les moyens choisis sont-ils les meilleurs possibles

Les moyens choisis aboutissent-ils à d'autres fins

b) fidélité à une méthode

c) choix de critères en fonction d'une classe. Exemple : juger Balzac sur des critères sociologiques, A. Dumas sur des critères historiques, A. Thierry sur des critères poétiques.

CHAPITRE VII

EXAMEN ET CONTROLE DES CONNAISSANCES

C - DEFINIR LES OBJECTIFS

Mihi à docto doctore
Domandatur causam et rationem quare
Opium facit dormire ?
A quoy respondeo :
Quia est in eo
Virtus dormitiva,
Cujus est natura
Sensus assoupire.

MOLIERE.

LE CONTROLE DES CONNAISSANCES ENVISAGE D'UN POINT DE VUE NOUVEAU 284
POUR OU CONTRE LE CONTROLE DES CONNAISSANCES ? 286
CONTROLE DES CAPACITES INTELLECTUELLES ELEMENTAIRES 288
FAUT-IL ALLER PLUS LOIN ? 293
CONCLUSION 296

LE CONTROLE DES CONNAISSANCES ENVISAGE D'UN POINT DE VUE NOUVEAU

Le grand succès de l'expression atteste d'une acceptation générale. Cependant, si "contrôle" même peut être pris dans plusieurs acceptions que dire de "connaissance" ? C'est un terme bien difficile à définir et, cependant, si l'examen est un "contrôle des connaissances", il est bon de savoir, au moins, ce qu'il contrôle. Dans l'acception commune, "connaître" et "savoir" sont synonymes : tout dictionnaire le confirme bien que cette synonymie ait quelque chose de gênant. "Sapere", c'est "avoir du goût"(de la saveur). Peut-être, cela suggère-t-il que, pour reconnaître un goût, il faut d'abord l'avoir connu et mis en mémoire. La reconnaissance proprement dite étant une comparaison entre des données immédiates et des données fournies par la mémoire. En fait, il semble que tous les auteurs s'accordent pour mettre en rapport "mémoire" et "connaissance". On peut considérer qu'une épreuve de connaissance suscite une manifestation de remémoration : la restitution d'informations, simples ou complexes, sous la forme même où elles ont été fournies(ou presque), sans manipulation. Ainsi se trouve définie, par un comportement(de remémoration), la connaissance. Ce n'est pas l'adaptation de l'activité à un but nouveau et précis, aucune manifestation d'intelligence, de raison, n'est demandée.

On peut juger qu'un tel objectif manque d'ambition. A cette objection, on peut répondre deux choses :

Premièrement : rien (ni personne) n'oblige à arrêter là, le niveau des exigences.

Deuxièmement : beaucoup d'examens se contentent de cet objectif qu'ils l'avouent ou non. Aucune différence de comportement n'est réellement exigée dans les deux questions suivantes :

1) Dates du commencement et de la fin du "Blocus continental".

2) Politique maritime de Napoléon 1er de 1806 à 1815

Répondre à la première question exige deux dates(mémoire) avec une précision qui peut aller de l'année au jour du mois. La seconde question, de rédaction plus longue, peut n'exiger qu'un compte rendu de lecture appris par coeur et n'est pas du tout un exercice garanti de synthèse.
Si le professeur et l'examinateur sont une seule et même personne, il sera possible de juger que le candidat a fait un exercice de mémoire(le professeur en avait parlé) ou un exercice de synthèse, une communication originale(le sujet n'avait jamais été traité). L'examinateur qui n'aurait pas été l'enseignant ne peut pas juger du genre de question... qu'il doit juger.

La vraie question est de savoir, si on accepte ou non l'expression "contrôle des connaissances" avec la notion de "connaissance mémoire pure" ou non. Si oui, l'opération est rigoureusement définie. Si non, on a le choix entre changer le sens du mot "connaissance" ou changer l'expression en précisant la nature de ce qu'on veut contrôler. Considérant que le vocabulaire est déjà suffisamment compliqué et que les opérations sont mal définies par des termes ambigus, il sera envisagé ici, le contrôle des connaissances et là, celui de la compréhension ou de l'application ou ... de la synthèse.

Ainsi deux grandes classes d'objectifs peuvent être définies :

les objectifs de mémoire(connaissance),
et tous les autres qu'on peut regrouper sous le terme de capacités intellectuelles.

POUR OU CONTRE LE CONTROLE DES CONNAISSANCES

En soulignant que "contrôle" est ici traduit par "vérification" et en rappelant que les "connaissances" sont des informations mémorisées, la question peut se transformer en "pour ou contre la vérification du contenu de la mémoire ?" Répondre à une telle question n'est pas aussi simple qu'il peut sembler : ni oui ni non ne sont satisfaisants.

Si l'examen a pour santion un certificat attestant que l'impétrant est informé, les connaissances doivent être contrôlées. Si l'information est le seul but, ce contrôle suffit ; si le but est différent, c'est autre chose qu'il faut vérifier : le contrôle des connaissances n'est pas toujours suffisant. Est-il toujours nécessaire ?

Si l'examen aboutit à une sanction : accepté ou refusé, il se pose la question de la justice d'un éventuel refus ou d'une éventuelle acceptation sur des connaissances jugées suffisantes ou non. Il semble bien que le critère de justice soit ici l'utilité et c'est là, source de discussions infinies. Rien ne peut être qualifié de totalement inutile. Il est tout à fait exceptionnel de rencontrer un sujet ayant appris le sanscrit qui déclare cette acquisition inutile. Cependant, si on veut bien séparer "l'humain" et la "culture" du "nécessaire pour l'accomplissement d'une tâche définie dont les études sont le but", le débat se simplifie : il est d'autant plus simple que le niveau des études est plus élevé, puisque la polarisation en est plus nette.

Ainsi, il est fort intéressant de savoir que scialytique et écureuil ont même étymologie, que l'animal est censé se mettre à l'ombre sous sa queue et que l'instrument fournit un éclairage sans ombre ; que le cubitus(os de l'avant-bras) est réputé être l'os du coude ; que l'épine dorsale a longtemps été prise pour un os unique, articulé, épineux et contenant, comme d'autres, de la moelle(la moelle épinière) ; que le nerf ischiatique a dû à sa plus grande célébrité de devenir sciatique, l'artère restant ischiatique...etc. Mais peut-on, honnêtement, refuser un candidat dans une épreuve d'anatomie, s'il ne sait pas expliquer l'origine du terme "moelle épinière" ? Cependant, la connaissance du vocabulaire est évidemment indispensable : sans elle que deviendrait le sens des discours, écrits ou parlés ? Beaucoup de gens vivent, apparemment, bien, sans "connaître "le théorème de Pythagore (I 31 : connaissance d'une loi) mais si on juge bon d'exiger de candidats qu'ils connaissent ce théorême, la moindre des choses ne serait-elle pas de s'assurer(vérifier) qu'ils connaissent, par exemple, le sens du mot "hypothénuse" ?

Trois inconvénients principaux sont attachés non au fait de contrôler les connaissances,mais à l'habitude qu'on risque de prendre(ou qui était déjà prise depuis fort longtemps) de le confondre avec examen :

1) S'en contenter à tous les niveaux de l'enseignement et, pour chacun d'eux, dans tous ses temps.

2) Attacher une importance directement proportionnelle au caractère abstrait, nouveau ou rébarbatif de la connaissance, ce qui est souvent le vice des spécialistes.

3) Négliger des connaissances par trop élémentaires par les enseignants qui jugent sur leur propre culture actuelle et non sur celle de la population qui leur est confiée. Le résultat est assez souvent monstrueux : c'est le psittacisme.

Tout jeune Français a la chance, pour peu qu'il fasse des études secondaires, de lire l'une des plus féroces critiques de l'enseignement : les ESSAIS de Montaigne. Et on ne peut que souscrire, quatre siècles plus tard, à certains jugements tout à fait applicables au contrôle des connaissances.

"La recherche des frases nouvelles et de mots peu cogneuz vient d'une ambition puérile et pédantesque... l'imitation du parler, par sa facilité, tout un peuple suit incontinent ; l'imitation du juger, de l'inventer ne va pas si vite... la force et les nerfs ne s'empruntent point ; les atours et le manteau s'empruntent". "C'est témoignage de crudité et d'indigestion que de regorger la viande comme on l'a avalée".

On ne sait de quoi s'émerveiller : que de telles choses aient été dites si tôt ou qu'elles n'aient pas encore été entendues si tard.

Au total, le contrôle des connaissances a sa place dans toutes les sortes d'études, et, vraisemblablement, à tous les niveaux mais on ne voit pas qu'il puisse être suffisant dans quelque partie que ce soit des enseignements. Toujours nécessaire, aussi longtemps que certains faits, nouveaux pour l'étudiant, sont abordés, il n'est suffisant qu'à un stade élémentaire. Il faut le combiner avec les autres contrôles d'une manière organisée qui constitue un "programme d'examen". Enfin, et dans cette perspective, il est indispensable, même si le coeur de beaucoup d'enseignants en saigne de distinguer, dans cet ordre de faits, ce qu'il faut connaître en permanence, ce qu'il est suffisant d'avoir connu et ce qu'il est totalement inutile de connaître : par exemple, le massage dans l'Antiquité tel qu'il figure au programme officiel des études de kinésithérapie.

CONTROLE DES CAPACITES INTELLECTUELLES ELEMENTAIRES

Il semble intéressant de grouper sous ce titre "compréhension" et "application". En termes d'examen, et que ce mot plaise ou non, ce qui est demandé comme preuve des connaissances, c'est une récitation ; sans faire de différence sur la longueur, dire combien font deux plus deux, donner la définition d'un mot ou exposer la théorie des quanta d'après Max PLANCK sont des récitations fondamentalement. Ce que beaucoup d'enseignants demandent à leurs élèves, c'est qu'ils "sachent réellement" ou qu'ils "comprennent". Pour certains, "savoir réellement" peut se traduire : "dire comme je l'ai dit". Pour d'autres, cela veut dire "comprendre". Les premiers demandent des connaissances, les seconds de la compréhension, mais il est tout à fait possible que ne soient pas compris ceux qui réclament la compréhension si "comprendre" n'est pas compris : pour les uns, comprendre, c'est "bien savoir", pour d'autres, c'est..."comprendre" ! Pour obscurcir encore un peu plus le langage, beaucoup d'étudiants déclarent qu'ils ne peuvent connaître que ce qu'ils comprennent. Pour saisir le sens de ce mot tel que le "comprend" BLOOM, les étudiants semblent être une excellente référence. Certains, aux cours, prennent des notes. Parmi ceux-ci, plusieurs catégories sont bien connues des enseignants.

Les uns "copient"... s'ils en ont le temps. Ils réclament de l'orateur,lorsqu'ils l'osent, qu'il parle moins vite, à la vitesse de la dictée, ou encore, ils se munissent d'un magnétophone. Le résultat est(ou tend à être) de la haute fidélité. A s'en fier aux seules notes, ces étudiants n'ont rien compris. Ils l'avouent d'ailleurs ingénument, en disant que, si fort occupés à noter, ils n'ont pas eu le temps d'écouter !

D'autres écrivent énormément, mais presque pas un mot du professeur n'est reproduit. Le sens y est parfaitement respecté, la forme a changé. Ces étudiants ont écrit avec leurs propres mots ce qu'ils entendirent : ils ont transposé. Ils ont compris.

D'autres encore écrivent peu. Paresseux organisés, ils savent reconnaître dans un exposé les artifices destinés à le rendre plus clair, plus vivant ; ils savent d'ailleurs en profiter, mais ils ne recueillent que l'essentiel, ils résument. Ils ont compris. Que font-ils ensuite ? Au cours suivant, ils prévoient souvent ce qui sera dit, si c'est une conséquence logique du précédent. Mais pour l'examen, ils auront tendance à "apprendre" leurs notes concises et à les "réciter" : ce jour-là ils feront preuve de connaissances.

Parmi les raisons qu'on peut trouver au fait qu'un étudiant se rapproche d'une étudiante, il en est une d'une affreuse vénalité : les notes féminines sont généralement mieux présentées, plus esthétiques et mieux écrites que les masculines. Ainsi un professeur juge, parfois avec raison, que son cours a été compris, mais il ne sait pas nécessairement par qui.

D'autres enfin ne prennent aucune note(certains, dit-on, ne viendraient même pas écouter le cours). Parmi ceux-là, certains sont doués d'une mémoire auditive ou "scénique" telle que la seule assistance à la performance du maître tient lieu de notes. Mais il en est qui utilisent les notes de leurs "copieurs" ou d'autres informations... qui seront éventuellement récitées. Pour qu'un examinateur croit réellement avoir testé la compréhension de ses candidats, il faut qu'il ait, lui-même, compris beaucoup de choses ou qu'il soit naïf... ou qu'il n'exige pas réellement la compréhension qu'il prétend exiger.

L'APPLICATION est un redoutable piège à examinateur. Les étudiants sont généralement enchantés à l'idée d'appliquer ce qu'ils savent, les enseignants en font volontiers un test d'efficacité et les examinateurs s'efforcent à trouver de bons sujets. En prenant le terme dans le sens qu'en propose BLOOM, ces questions sont un réel test d'efficacité, mais les candidats ne les aiment pas beaucoup et les examinateurs se font beaucoup d'illusions en croyant qu'il est simple de faire de telles questions. Il ne s'agit absolument pas d'une question concrète, mais du résultat du libre choix de l'abstraction appropriée.

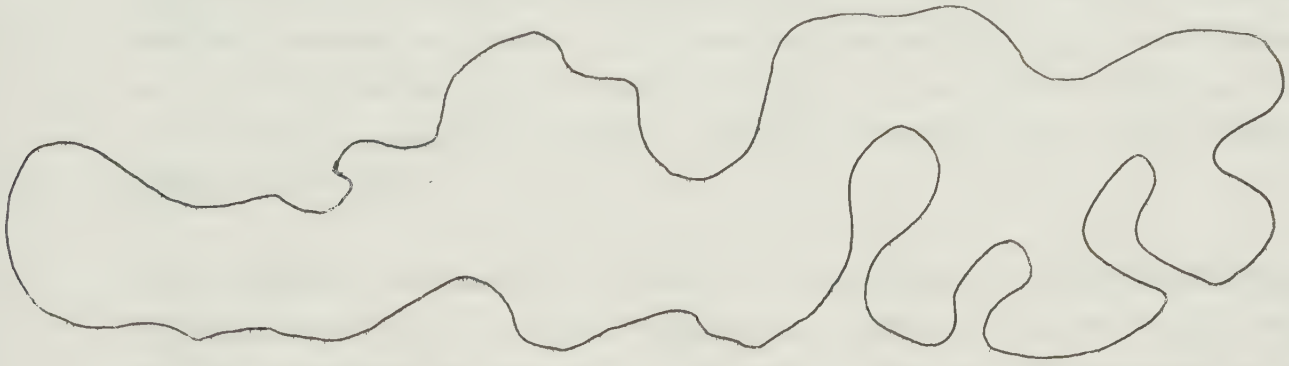
Ainsi, une question du genre "une pomme vaut dix centimes, combien valent sept pommes" ?, n'est pas une application, mais un exercice de connaissance des moyens d'utilisation des informations élémentaires (I20) et, plus précisément de technique(I 25) : ce n'est pas une abstraction qui est utilisée, mais un "truc". Il en est de même de beaucoup de problèmes pour la résolution desquels il suffit d'utiliser(on dit, à tort, appliquer) une formule qui n'est même pas nécessairement comprise(1)En fait, appliquer c'est choisir : un exercice d'application doit donc laisser le choix. Si l'abstraction utilisée est imposée, c'est un exercice de compréhension.

(1) cf. "une question, 17 + 1 réponses" p. 144 -

Toute la difficulté réside dans le discernement nécessaire entre le libre choix apparent et qui est, en fait, imposé et le libre choix réel.

Voici un exemple :

- Question : quelle est la surface de cette figure à 1 p. 100 près ?



Trois bonnes réponses sont possibles dont : "je ne sais pas". Les voici, présentées avec un cynisme dont peu de candidats ont l'audace.

1) Je décèle immédiatement que, cette figure étant irrégulière, aucune formule(abstraction) ne peut me servir ; j'estime qu'en "rectifier" les contours entraînerait une erreur supérieure à vos exigences.Je vous en informe et déclare le problème sans solution.

Si vous, mon professeur ni aucun de vos prédécesseurs, ne m'avez jamais appris à résoudre cette sorte de problème, j'ai montré ma capacité(et même mon aptitude)... à appliquer. Mon comportement est exactement celui du schéma de BLOOM cité en fin de chapitre VI.

a) aucune restructuration ne peut m'amener à un type connu me permettant le choix d'une abstraction(application).

b) aucune abstraction ne pouvant, par moi, être retenue, je ne peux en appliquer aucune(solution ; la difficulté du vocabulaire consiste à ne pas confondre choix de l'abstraction et utilisation de l'abstraction ou application et solution).

c) la seule réponse correcte que je puisse faire est "je ne sais pas"(réponse)

Et j'espère bien avoir la meilleure note possible, sauf si vous êtes certain que j'ai pu répondre autre chose.

2) Je me trouve devant une situation nouvelle(à vous de savoir si je dis vrai). Je pense que les n centimètres carrés de la figure découpée pèseront n fois plus qu'un centimètre carré : c'est (a) le choix de l'abstraction.

Le reste est extrêmement simple. Tellement simple que j'espère bien que votre exigence d'une précision de 99 p. 100 n'est qu'un leurre destiné à ne pas me souffler la réponse ou à me mettre sur la voie. Exiger que j'aille plus loin n'est justifiable que si vous voulez tester ma capacité à utiliser une paire de ciseaux et une balance de précision. Mais, est-ce réellement le but ?

3) Je connais parfaitement l'existence d'une méthode "d'intégration physique" ou "graphique" et la solution s'impose. Je ne fais pas preuve d'application mais de compréhension puisque je n'ai pas le choix de l'abstraction qui s'impose.

De cet exemple, il ressort que deux candidats ont fait un réel exercice d'application, dont l'un n'a pas fourni la réponse exigée, et un autre un exercice d'un niveau moins élevé. La réponse exigée peut être considérée comme inutile mais l'exigence qu'on en manifeste présente une certaine utilité. Enfin, et surtout, il ne suffit pas d'avoir conçu la question pour savoir quel est son niveau, il faut encore connaître le niveau réel des candidats auxquels elle est destinée.

Au total, la moindre des choses qu'on puisse exiger d'un enseignement, c'est qu'il fournisse des informations(des connaissances) mais ce n'est pas suffisant. Incidemment, on peut remarquer que quantités de matériels semblent pouvoir donner ces informations aussi bien et peut-être mieux qu'un professeur. Au-delà des connaissances, il est toujours indispensable d'aller au moins jusqu'à l'application ce qui ne signifie pas "utilitaire" ou concret.

Dans beaucoup de domaines, ce qui est entendu par le mot "application" est en fait une utilisation de techniques qui suffit à rendre efficace l'ex-étudiant ou l'ex-apprenti, mais cette efficacité immédiate est un leurre. Considérée comme une fin en soi, une certaine technicité est stérile car il est bon de se convaincre que les modifications incessantes qu'on appelle en général "progrès" sont autre chose qu'un cliché seulement utile à remplir les lignes d'un journaliste besogneux.

Si la technique, le "truc" actuel est souhaité(et souhaitable), il est nécessaire qu'il soit atteint, notamment, par une voie de facilitation(et non de facilité) qui est l'application : le choix. D'autre part, s'il est parfaitement vain de soutenir qu'on peut être instruit sans mémoire et savant sans connaissance, il est tout aussi vain de nier que l'exercice des capacités intellectuelles est sans effet sur l'acquisition de nouvelles connaissances. Pour une discipline donnée, l'enseignement doit dépasser le niveau des simples connaissances. S'il est démontré que des connaissances suffisent, il faut remplacer l'enseignant par un guide bien rédigé, le mot "discipline" par un autre et se contenter d'un contrôle des connaissances.

FAUT-IL ALLER PLUS LOIN

On peut comparer analyse et compréhension, synthèse et application. Analyse et synthèse sont des niveaux d'activité supérieurs dont on peut se demander, pour une discipline donnée s'ils sont réellement utiles.

Exercer une analyse sur un certain sujet, c'est, en quelque sorte, entrer dans son intimité et la question est de savoir si cette pénétration apporte quelque chose d'indispensable. Si l'activité souhaitée d'un étudiant est la recherche des éléments, des relations entre eux ou des principes de leur organisation, l'analyse est un objectif. Par exemple, l'interprétation d'un document radiologique pour un étudiant en médecine est une analyse dont la réalisation est un besoin quotidien dans l'exercice. De même, un histogramme pour un statisticien, du compte-rendu d'une expérience pour un physiologiste... etc. Dans ces conditions, l'atteinte de l'objectif doit être contrôlée.

L'analyse peut être une étape indispensable à la compréhension de faits complexes. Si ces faits doivent être compris, la compréhension est un objectif d'examen, l'analyse un moyen d'enseignement qui n'a pas à faire l'objet d'un contrôle. Ainsi, au cours des études de médecine, en physiologie par exemple, et plus généralement dans les disciplines dites fondamentales, l'analyse est un moyen utile pour enseigner mais ce n'est pas un objectif d'examen pour des gens pour qui la science considérée n'est, finalement, qu'une série d'information Cependant, on peut estimer que l'analyse de données physiologiques soit utile si on la considère comme un exercice d'entraînement relativement simple sur un sujet qui présente des analogies avec une autre sorte de situations qu'il sera réellement indispensable de savoir analyser. Dans ces conditions ce n'est plus envisager une discipline sous son aspect technique(de spécialiste) mais éducatif : fournissant un certain modèle utilisable ailleurs.C'est une étape.

Dans la mesure où l'objectif le plus élevé est la critique, il est indispensable de contrôler la capacité d'analyse qui est, ici aussi, une étape. Contrairement au cas précédent où l'étape pouvait être qualifiée de facilitatrice, il s'agit ici d'un intermédiaire obligatoire. Une mauvaise critique peut être la conséquence d'une mauvaise analyse et une mauvaise analyse doit logiquement aboutir à une mauvaise critique. Face à une mauvaise critique, il ne sera pas aisé de discerner où se situe la faute, à quel stade de la démarche, si l'épreuve elle-même n'a pas été constituée de manière analytique.

Ainsi, l'analyse semble un objectif d'enseignement indispensable dès qu'il est requis de comprendre. C'est un objectif d'examen(l'analyse doit être l'objet d'un contrôle) dans tous les cas où il est requis de l'étudiant qu'il dépasse le simple niveau des informations acquises et comprises. Certaines disciplines qui peuvent en réalité n'être considérées qu'en tant que sources d'informations peuvent faire l'objet d'investigations d'un niveau plus élevé(donc d'exigences supérieures) dans la mesure où leurs propriétés "éducatives" sont retenues sans tomber dans l'excès fréquent consistant pour les enseignants, à considérer tout étudiant comme un futur spécialiste de sa propre discipline.

X

X X

De la synthèse, que certains n'hésitent pas à confondre avec une certaine créativité, il faut beaucoup se méfier. Il est certain qu'une leçon est un exercice de synthèse pour l'enseignant et, à fortiori, l'ensemble des leçons constituant un cours. Comme dit Montaigne, "l'imitation du parler, par sa facilité, tout un peuple suit incontinent ..." Beaucoup d'étudiants souhaitent faire de leur "copie" une leçon magistrale et beaucoup d'enseignants les encouragent dans cette voie, mais la leçon magistrale qu'ils souhaitent lire, c'est la leur propre : ce n'est pas de la synthèse mais du psittacisme.

Si on veut bien se rappeler que la condition nécessaire à la réalisation d'une synthèse est la liberté de pensée et d'action, sans point de vue ni conclusion imposés, on se rend aisément compte qu'il s'agit là d'un exercice très différent d'une récitation. Un diagnostic, un traitement à prescrire ou tout autre plan d'action à élaborer sont des synthèses. Il est évident que tout individu libre agit consécutivement à l'élaboration de son propre plan d'action : la synthèse semble ainsi une condition nécessaire à la liberté et tout individu possède, à un niveau plus ou moins élevé, une telle aptitude que l'enseignement de l'enfant ou de l'adolescent doit certainement favoriser.

Lorsqu'on en arrive à un stade très orienté des études, c'est dans certains domaines seulement que doit être exigée la capacité de synthèse. Ainsi, pour un médecin, exiger (1) de lui qu'il élabore un plan d'action dans une discipline non strictement médicale, telle la physiologie, est un non sens : c'est une perte de temps si l'objectif n'est pas une spécialisation et une caricature si la formation préalable(le niveau des exigences) n'a pas été suffisante.

Inversement, la synthèse n'est pas un domaine réservé aux "savants" aux "purs intellectuels" : elle n'est pas seulement la constitution d'un ensemble de relations abstraites, elle peut être aussi l'élaboration d'un plan d'action et peu de professions, à un degré ou à un autre, ne requièrent pas, pour leur exercice, une telle opération. Enfin, l'exercice le plus simple, celui dont l'utilité est la plus communément admise : la communication originale dans un but précis, la rédaction d'une lettre, par exemple, ne semble pas faire l'objet de beaucoup d'exigences quand il faudrait : tôt (primaire et secondaire) et ceci expliquerait peut-être une partie du réel désarroi des études supérieures par l'habitude qu'on prise les étudiants de dire n'importe quoi, n'importe comment, scandalisés de n'être pas compris.

Enfin, il n'est pas de synthèse possible sans critique et ceci est particulièrement visible dans l'élaboration d'un plan d'action : la moindre des choses est de vérifier que la démarche entreprise est en rapport avec le but souhaité. Toute vérification est par ailleurs une critique. Par exemple, l'opération consistant à faire passer un examen nécessite une critique interne(rigueur, cohérence, justesse...) aussi bien qu'externe : l'épreuve proposée est-elle réellement probante de ce qu'on veut prouver ?

(1) ne pas confondre exiger et permettre.

CONCLUSION

Au total, il peut être utile, voire nécessaire d'aller plus loin que l'exercice de capacités intellectuelles élémentaires.
Pour un examen, la question est justement de discerner l'utile du nécessaire, de retenir le nécessaire aux fins de vérification. Il est bien évident que l'esprit apte aux vastes synthèses aura beaucoup plus de facilités pour "apprendre" c'est-à-dire mémoriser des informations et notamment les ensembles conceptuels. Il faut bien, à un certain moment, se décider à vérifier que le candidat possède réellement ce qu'il sera réputé posséder lorsqu'il sera reçu et savoir se contenter d'un contrôle des connaissances là où cette seule acquisition est requise et savoir exiger un niveau beaucoup plus élevé là où il est reconnu indispensable qu'il soit atteint. Par exemple, se contenter d'une récitation là où elle est suffisante et l'empêcher par tous les moyens là où, insuffisante, elle pourrait abuser l'examinateur sur ce que montre réellement le candidat de ses capacités. C'est-à-dire qu'il est très utile de discerner objectifs d'examen et moyens d'enseignement.

CHAPITRE VIII

OBJECTIFS DEFINIS D'UN EXAMEN

C - DEFINIR LES OBJECTIFS

La définition des objectifs d'examen aboutit à une liste de comportements décrits, observables et vérifiables permettant d'affirmer ou de nier que le but de l'enseignement est atteint. Celà constitue en quelque sorte, une règle du jeu qui risque de diminuer sensiblement le caractère despotique de l'épreuve.

Par ailleurs, une telle opération permet de confondre, ou de tendre à confondre, le but des études, le but de l'enseignement et le but du travail fourni par l'étudiant: ainsi peut-on espérer que la qualité du produit fini s'améliore surtout si on considère comme une qualité le fait que le diplômé possède réellement ce que le diplôme garantit.

DEFINITION D'UN OBJECTIF D'EXAMEN 298
DEFINITION D'UN OBJECTIF D'ENSEIGNEMENT 299
COMBIEN FAUT-IL DEFINIR D'OBJECTIFS ? 300
COMMENT DEFINIR LES OBJECTIFS ? 301
UTILITE DE LA DEFINITION DES OBJECTIFS 305

DEFINITION D'UN OBJECTIF D'EXAMEN

Définir un objectif d'examen, c'est le décrire en termes tels que tout le monde le comprenne de la même façon.

Une telle démarche revient à faire part de ses intentions, en termes clairs : c'est une communication. Il n'est pas possible de se contenter de proclamer qu'il sera exigé de "connaître", de "savoir", d' "avoir compris", de "réellement savoir", de "juger"....etc., car aucun des mots clés d'une telle définition ne peut être prévu comme représentant les mêmes choses pour différents lecteurs. En supposant même que des termes fondamentaux, tels : connaître, comprendre, appliquer... aient été eux-mêmes définis de telle sorte qu'aucune ambiguïté ne soit possible, leur utilisation pour la description d'un objectif d'examen resterait insuffisante, car ne serait pas précisée la nature de la preuve attendue.

Soit une épreuve d'examen ayant pour thème... l'examen. Le sujet choisi peut être ainsi défini : "connaissance du vocabulaire". Cette définition implique que l'épreuve porte uniquement sur la connaissance des mots et que, d'une manière ou d'une autre, elle sera constituée de définitions : soit de mots(sujets) à définir(réponses), soit de définitions(sujet) à identifier (réponse). Il est exclu que, par exemple, soit demandées sous une forme ou une autre, les similitudes et les différences entre programme d'examen, programme électoral ou programme de la fête : ce ne serait pas une épreuve de connaissance. Même en supposant que soit fournie une liste limitative de mots à connaître, l'objectif d'examen serait insuffisamment décrit, insuffisamment précis : faut-il fournir la preuve de sa connaissance sur tous les mots de la liste ? sur tous les mots proposés par l'épreuve ou sur une certaine fraction seulement ? Quelle fraction ?

Sera-t-il demandé de passer du mot à sa définition ou de la définition au mot décrit ? On peut penser que les deux sens sont aussi utiles : l'un permettant de savoir lire et l'autre de savoir écrire : il faut le dire. L'objectif peut alors se définir ainsi :

"Etre capable d'identifier la définition des mots habituellement utilisés en docimologie ou de donner la définition correcte de ces mots dans une proportion d'au moins de quatre-vingts pour cent"

ou encore : "Etre capable d'identifier la définition correcte de ces mots. Dix-sept réponses correctes seront exigées sur une liste proposée de vingt".

Dans la deuxième version, l'épreuve même est définie dans sa forme et ses critères de réussite.

DEFINITION D'UN OBJECTIF D'ENSEIGNEMENT

Définir un objectif d'enseignement, c'est le décrire en termes tels que chacun puisse le comprendre de la même façon.

Une telle démarche revient à faire part de ses intentions en termes clairs : c'est une communication. Avant de décider que le cours "fera connaître", expliquera pour aider à "comprendre", il est nécessaire que ces mots représentent des activités précises et c'est à ce niveau que risquent de se glisser de graves erreurs. Qu'on veuille bien imaginer qu'un enseignant soit chargé d'enseigner à ses collègues comment concevoir un examen. Il sera tout à fait louable que son ambition soit de leur faire connaître ce qu'ils ne connaissent pas et de faire en sorte que son auditoire comprenne certaines choses plus vite que lui-même n'y est parvenu : voici des objectifs que personne ne contestera. Si, devant une commission, celui qui est chargé de cet enseignement interne propose de diffuser des textes, de prononcer des conférences ou de présenter des films, on peut prévoir que peu d'objections lui seront opposées : ce sont là des moyens habituels.

Il n'en serait vraisemblablement pas de même si ses projets consistent à chanter, à danser, ou à diriger une représentation dramatique car ces moyens d'enseignement ne sont pas habituels et ne feraient pas "sérieux". Peut-être, quelques membres de la commission le loueraient-ils de son audace et le soutiendraient-ils pour son anticonformisme. Si un vote d'approbation devait survenir, il ne serait pas inimaginable que les deux partis ne puissent se départager. Qu'on appelle un arbitre, comment jugerait-il, s'il n'est ni conformiste, ni conformiste de l'anticonformisme ? Sans doute en posant la question qui n'a jamais été soulevée : comment savoir que le but proposé aura vraiment été atteint ? Si l'expérience montre qu'il est plus efficace de chanter son cours que de le dire, il faut le chanter. Cela suppose que l'objectif ait été défini en termes tels qu'on puisse savoir s'il a été atteint : qu'il ait été défini en termes d'examen(de vérification). La commission aurait été divisée sur une question de moindre importance, d'importance secondaire. Sa faute aurait été de confondre objectif et méthode. Son excuse aurait été de suivre de nombreux exemples.

Un objectif d'enseignement n'est pas défini si sa définition ne précise pas les critères permettant d'affirmer ou de nier qu'il a été atteint. Si on considère l'enseignement comme autre chose qu'une grille d'emploi du temps, des lieux et des personnes, ses objectifs doivent se confondre avec ceux de l'examen. Définir le but de l'enseignement c'est définir les preuves permettant d'affirmer que l'objectif a été atteint ou non.
C'est l'objectif de l'examen qui doit être défini.

COMBIEN FAUT-IL DEFINIR D'OBJECTIFS ?

Autant qu'il est nécessaire pour faire part clairement de ses intentions, c'est-à-dire, souvent, beaucoup.

Soit par exemple, la radiographie qu'il s'agit d'enseigner aux étudiants en médecine et, plus précisément, dans le cadre de la rhumatologie. "Connaître les éléments de radiologie indispensables en rhumatologie" est une définition, certes plus précise que "connaître les éléments de radiologie utiles". La seconde définition dans laquelle le mot clé est "utile" sera comprise de façon nécessairement différente par le physicien, le traumatologue, le neurologue ou le gastro-entérologue. La première suppose déjà une certaine communauté d'intérêts entre gens ayant la même activité et implique qu'elle soit répétée, en termes comparables, dans les différentes disciplines.

Soit, ensuite, à discerner ce qui est indispensable : tout ce qui sera reconnu mériter ce qualificatif sera retenu comme objectif... qu'il ne restera plus qu'à définir. En effet, le mot clé : "indispensable" représente une déclaration d'intentions(limiter les exigences), mais pas une liste d'objectifs précis. Si une telle déclaration est présentée à des étudiants, ignorants par définition, ils interprètent eux-mêmes, constituent par eux-mêmes la liste souhaitée en s'exposant aux plus grossières erreurs puisqu'ils sont ignorants. La seule chance qu'ils aient de ne pas se fourvoyer est de tenter de déceler, à forces d'observations patientes, ce que leur Professeur estime indispensable. Finalement, dans le meilleur des cas, où le Professeur est l'examinateur, le résultat dépendra de facultés de discernement qui n'ont rien de radiologique. En admettant même que les rapports entre professeur et étudiants aient été assez fréquents et prolongés pour que l'enquête ait disposé d'un temps raisonnable, on peut supposer que le temps eût été mieux employé à découvrir autre chose que les marottes du Maître. Dans le plus mauvais des cas, l'examinateur n'ayant pas eu l'occasion de laisser deviner ses marottes, les étudiants en seront pour leurs frais(sauf ceux dont l'examinateur était aussi l'enseignant). Dans les deux cas, les résultats seraient injustes, puisque le critère de jugement réel serait la capacité de discerner ce qui est indispensable, c'est-à-dire autre chose que ce qui avait été annoncé.

COMMENT DEFINIR LES OBJECTIFS ?

D'une telle manière que les étudiants sachent ce que l'on attend d'eux, et que les enseignants sachent ce qu'ils attendent de leurs étudiants; qu'il en résulte, pour les uns comme pour les autres, une organisation de leurs activités et que l'énergie dépensée le soit dans un sens qui corresponde à la finalité des études entreprises. Les guides à choisir pourront donc être l'utilité et l'efficacité.

L'exemple choisi sera la radiologie telle qu'on peut la concevoir dans le curriculum normal des études de médecine, et non telle qu'il serait bon qu'elle soit conçue en vue d'études de spécialité. Trois questions seront envisagées : utilité de la radiologie
: que faut-il en savoir
: et comment déterminer que ce qui doit être su l'est.

Dans le cadre d'études choisi, on peut avancer que la radio aide au diagnostic(dans une proportion variable selon les cas). On retiendra donc comme activité utile ce qui aide au diagnostic : on ne se contentera pas de faire savoir l'utilité de la radio mais on exigera que l'étudiant sache utiliser la radio puisque les conditions prévisibles d'exercice ne l'amèneront pas à faire un discours ou à écrire un article sur l'utilité mais à utiliser lui-même.

"Utiliser" n'est pas très clair : le choix fait n'est pas suffisamment précis. Faut-il savoir prescrire, savoir exécuter ou savoir interpréter. Savoir interpréter(lire) est incontestablement indispensable puisqu'il s'agit de recueillir d'un document ce qu'il peut fournir des renseignements qu'on lui demande. "Incontestablement" est peut-être un peu fort. On peut supposer quele document soit lu par celui qui le fait et considérer comme normal qu'il en soit ainsi. Dans ce cas, savoir interpréter est inutile et ce qu'il reste à savoir est l'utilité de la radio et non son utilisation. Un tel choix ne peut être fait à l'échelon de la faculté, sauf si cet échelon définit en même temps une circonscription d'exercice. Si la circonscription d'exercice est nationale, le choix doit être national ou international, s'il est possible d'exercer au-delà des frontières.

En retenant que, en France et dans l'état actuel des choses, il faut savoir interpréter, en admettant que cet objectif(mal défini) n'est ni universel ni éternel, faut-il savoir prescrire ? Il semble bien qu'on soit obligé de répondre "oui" sauf acceptation de changer complètement les conditions réelles d'exercice. Les conditions réelles sont ce qu'elles sont ; on peut souhaiter qu'elles changent, mais elles n'ont pas changé.

De toutes façons, rédiger un simple "bon pour un examen radiologique" suppose que soit connue l'utilité de la radio : c'est une prescription.

Une prescription plus précise telle "radios debout et couché, de face, de profil et de 3/4 droit, centrées sur L5-S1" suppose que soit reconnue aux deux positions et aux trois incidences une utilité particulière. Si le document est lu par le prescripteur, il lui faudra reconnaître si le centrage et les incidences sont bien ceux qui avaient été demandés. Savoir si la prescription a été bien exécutée : c'est, jusqu'à un certain point, savoir exécuter. Le terme "jusqu'à un certain point" mérite évidemment d'être précisé.

Décrire ou énumérer ce qu'il faut savoir de la radio est assurément plus clair que de se contenter de proclamer "il faut savoir un certain nombre de choses en radiologie". C'est ici que peut être utilisée une classification taxinomique simplement dans le but de ne rien oublier.

Exemple :

I 10 : connaissance d'informations élémentaires
11 : Terminologie
12 : Faits

Si l'enseignement de la radiologie est une partie de l'enseignement d'une discipline, le choix des mots sera plus restreint et leur sens plus précis que dans le cas où l'enseignement se voudrait "global". Par exemple, en rhumatologie, des termes tels que "becs de perroquet" ou "syndesmophytes" doivent être connus et ces faits identifiés sur un document.

Il n'est pas absolument nécessaire de fixer des objectifs dans chacune des catégories de la classification. Ainsi, et toujours dans la perspective de la rhumatologie, il ne semble pas que(I 32) la connaissance de la théorie de la radiologie soit exigible ni que cet enseignement trouverait là sa meilleure place possible.

Il n'est pas suffisant d'énumérer ce qu'il faut savoir, il faut encore définir ce que signifie "savoir" pour chaque objectif retenu.

Voici deux exemples :

1er EXEMPLE :

OBJECTIF DEFINI : connaître les différentes incidences utilisées dans les radiographies du genou et, pour chacune d'elles, connaître leur utilité particulière.

QUESTION D'EXAMEN : Après avoir examiné vous-même tel malade et plus particulièrement son genou gauche, demandez par écrit les radios qui vous semblent utiles.

Voici une question sans rapport avec l'objectif défini et deux méthodes permettent de le démontrer :

- la plus courte consiste à mettre en évidence la situation respective de l'objectif et de la question dans la classification taxinomique. L'Objectif décrit est une connaissance(I), une connaissance des moyens d'utilisation des informations élémentaires(I20) et, plus précisément, une connaissance de techniques(I 25). La question exige le choix d'une méthode(abstraction) : c'est une application(III). Il n'y a pas de rapport entre le but décrit et la question posée.

- la seconde méthode consiste en une analyse du contenu des deux propositions. L'objectif est décrit en termes de connaissances : elle implique, pour que le candidat soit jugé satisfaisant, qu'il restitue les informations reçues sans "manipulation" sensible. Pour chaque incidence qu'on peut utiliser(utilisable) il lui est demandé de dire ce qu'elle peut montrer comme détails anatomiques particuliers. Ni implicitement, ni explicitement, il n'est donné comme activité souhaitable, que soit jugée des indications particulières de chaque incidence fondée sur une entrée pathologique, ni surtout, choisie une incidence ou une série d'incidences à partir d'un cas particulier dont le candidat aura lui-même à juger. Cependant, c'est bien ce qui lui est demandé : examiner un malade, établir une probalité de diagnostic et choisir la technique radiologique la mieux appropriée à la confirmation. L'examinateur peut, dans l'absolu et à bon droit, refuser ce candidat s'il n'a pas demandé, par exemple, une radio pulmonaire, alors que le candidat, à bon droit, pouvait penser qu'il ne serait jugé que sur une prescription concernant le genou. Tout au plus, peut-il s'attendre à ce que des radios des deux genoux soient utiles pour l'épreuve.

2ème EXEMPLE :

OBJECTIF : Etre capable de lire une radiographie osseuse concernant n'importe quelle partie du squelette sous n'importe quelle incidence, chez un sujet ayant achevé sa croissance osseuse.

QUESTION D'EXAMEN : Après examen attentif du document qui vous est confié, décrivez-le : incidence utilisée, anomalies observées, diagnostic proposé.

La question d'examen demande une description et un diagnostic dont il semble admissible qu'il ne soit pas exact ou pas complet, vraisemblablement parce que le document est unique : c'est une question d'analyse suggérée par le "diagnostic proposé" qu'on peut traduire par (IV30) recherche des principes d'organisation. On peut espérer qu'il s'agit d'une radiographie osseuse. Cependant, les rapports entre l'objectif et la question restent ambigus. Alors qu'on demande de décrire(ce qui peut passer pour une synthèse V 10) il était fixé comme objectif de savoir lire et on peut reprocher à cette formulation de n'avoir pas précisé quels critères seraient utilisés pour juger que le candidat savait lire. Etant donné le genre de question posée(qui semble justifiable) il fallait décrire l'objectif autrement et dire (pourquoi pas ?) "savoir décrire"... ce qui ne changeait rien au niveau des exigences qui reste fondamentalement celui d'une analyse, mais simplifiait beaucoup le travail des candidats sans, pour autant, dénaturer le but.

UTILITE DE LA DEFINITION DES OBJECTIFS

L'enseignement représente pour les enseignants et les étudiants une activité mobilisant beaucoup de temps et de moyens. Sans but, cette activité n'est qu'agitation, l'enseignement n'est qu'une manie. Définir le but en termes d'examen n'est pas la manifestation d'un goût bizarre, une manie de l'examen(kakorrhaphiomanie).

Si l'enseignement(pour les uns) les études(pour les autres) ont une finalité, il semble utile de la connaître. Savoir qu'un but existe est une chose, tenter de l'atteindre, une autre. Savoir prouver qu'il est atteint est tout à fait indispensable sans quoi n'importe quoi peut être pris comme preuve de n'importe quoi. Si la finalité des études est l'acquisition de quelque chose de plus, ce qu'aucun enseignant ne nie, semble-t-il, la trace observable, la manifestation du résultat est l'apparition de capacités nouvelles qui peuvent être décrites en termes de comportement. Devenir capable de faire quelque chose de nouveau, c'est le but des études et montrer qu'on est devenu capable, celui de l'examen. Les objectifs d'examen ne sont rien d'autre que la description des nouveaux comportements à acquérir, c'est-à-dire les objectifs mêmes de l'enseignement. Confondre objectif d'enseignement et objectif d'examen suppose que la question est rédigée après que l'examinateur y aura lui-même répondu, que c'est en fonction de la réponse prévue que sera jugée la valeur de la question envisagée.

C'est encore une réalité que de considérer que l'acquisition de ce qui constitue l'ambition des études est un ensemble parfois fort complexe, qu'il est utile de savoir décomposer en éléments clairement définis et logiquement classés. Le classement logique c'est la taxinomie qui n'est rien d'autre qu'un système de classement permettant, au moins, d'approcher d'une liste exhaustive des objectifs à atteindre. Quand on sait ce qu'il faut faire, on peut tenter de savoir comment le faire. La définition des objectifs n'est qu'une étape, leur classement, un simple outil commode pour ne rien oublier. Il reste encore à les distribuer selon le temps, à les organiser, les uns par rapport aux autres, dans un ordre efficace : ceci est une programmation.

D - ETABLIR UN PROGRAMME

Il est malheureusement très facile de trouver dans les textes officiels des modèles de ce qu'il vaudrait mieux éviter. Voici quelques exemples illustrant le mot " programme "

. PROGRAMME D'ETUDES (CES de Biologie appliquée à l'éducation physique et aux sports. Annexe à l'arrêté du 15-12-1949).

Sous ce titre on trouve énumérées en trois chapitres: éducation physique, sport et rééducation physique, cinquante trois rubriques

. PROGRAMMES DES CONFERENCES (CES de Dermato-Vénéréologie; annexe à l'arrêté du 11-10-1949).

Le contenu de cent cinquante deux conférences s'y trouve énuméré. Chacune devant durer une heure et les études s'étendant sur trois ans, trente séances de travaux pratiques de 2 à 3 heures devant s'y ajouter, sachant que l'année dure 9 mois, on peut calculer que l'emploi du temps quotidien moyen est un peu inférieur à trente minutes.

. PROGRAMME DES CONNAISSANCES EXIGEES (CES de Cardiologie; annexe à l'arrêté du 7-09-1965). Examen de première année.

Se trouve énumérée, sans autres commentaires, une liste de 75 maladies et commencent par " Scarlatine, Rougeole, diphtérie, tétanos ,...." sur lesquelles seront jugés après un an d'études de spécialité des étudiants étant déjà docteurs en Mèdecine ou ayant terminé leurs études de Mèdecine. Celà en dit long sur la confiance accordée aux examens préalables !

On y trouve aussi de l'imprécis: la femme et les sports, les sports de base, les efforts sportifs .... et même du pittoresque: les sports individuels cités sont la boxe, l'escrime et le tennis; la formation prémilitaire figure dans le chapitre rééducation; on demande l'étude des mécanogrammes cardiaques artificiels (sic) et veineux...etc.

Rien de tout celà ne mérite le nom de programme:chacune des rubriques est rédigée de façon beaucoup trop vague.Programmes d'objectifs d'études ou d'examen sont choses différentes contrairement à ce que laiss supposer l'identité des formes précédentes.

CHAPITRE IX : PROGRAMMES

CHAPITRE X : CIRCONSTANCES D'UN EXAMEN

CHAPITRE IX

PROGRAMMES

D - ETABLIR UN PROGRAMME

Un programme d'objectifs est un catalogue précis et complet. Un programme d'enseignement est une distribution dans le temps, aussi logique que possible, des objectifs et l'organisation des moyens propres à y parvenir.
Un programme d'examen est une répartition des contrôles déterminée par le programme d'enseignement. De nombreuses variantes sont possibles en ce qui concerne le programme d'enseignement (donc celui des examens qui en dépend). Quant aux programmes d'objectifs, pour une matière donnée, ils ne peuvent guère se distinguer que sur le degré de leur précision et de leur achèvement.

Quant à la publication des objectifs, elle permet aux étudiants d'organiser , s'ils en ont envie, leur propre programme de travail.

PROGRAMME D'OBJECTIFS 308
PROGRAMME D'ENSEIGNEMENT 312
. Enseignement traditionnel 313
. Enseignement par blocs 315
. Enseignement intégré 322
PROGRAMME D'EXAMENS 326
PUBLICATION DES OBJECTIFS 328

PROGRAMME D'OBJECTIFS

Ce programme n'est qu'un catalogue : il fixe les buts à atteindre par les étudiants et par conséquent, ceux qui doivent être atteints par l'enseignement. Il constitue la liste de ce qui est exigible, étant bien entendu que ce qui sera exigé réellement risque d'être sensiblement plus réduit. Dans la mesure où les examinateurs et les enseignants ne sont pas confondus, le catalogue constitué ne peut être considéré que comme un minimum : ce doit être une liste de points communs entre tous ceux qui obtiendront finalement la même estampille. Par exemple, il arrive que dans une même faculté et pour une même année des mêmes études, les étudiants soient divisés en sections pour des raisons matérielles diverses le plus souvent tout à fait justifiées. Pour une matière donnée, deux sections peuvent être enseignées, chacune par un professeur. Lorsque l'examen approche, il peut se trouver que soit déclaré qu'aucune question commune aux deux sections ne peut être posée, l'enseignement ayant été trop différent. Voici pourtant deux catégories de candidats qui recevront la même estampille garantissant, en principe, la même chose.

On peut soutenir qu'il est difficile, impossible, inhumain, d'imposer à deux professeurs de "mécaniser" leur enseignement à un point tel qu'ils soient parfaitement interchangeables à n'importe quel moment et il n'est pas question de réfuter cet argument parce qu'il n'est pas sérieux. En fait, il faut choisir entre deux conceptions du "cours" : aider des étudiants à atteindre un certain but ou se donner le plaisir d'une performance éventuellement brillante, sur un sujet particulièrement intéressant pour l'orateur. Il semble bien que, jusqu'au second cycle des études supérieures inclus, la première conception soit meilleure que la seconde. Par ailleurs, imposer à deux professeurs les mêmes buts ne signifie pas que l'obligation porte aussi sur les moyens. Compte tenu de différences de personnalités, de styles, il est légitime que deux enseignements soient différents et refuser ce point serait être utopique. En revanche, il est scandaleux de constater l'absence de points communs, de tous points communs, dans des enseignements destinés par exemple, à des étudiants en médecine dans deux sections de la même faculté, s'adressant à la même année, car deux seules explications logiques peuvent être fournies. Ou bien chacun n'enseigne qu'une partie de ce qui est utile et il manquera quelque chose d'utile à tous les étudiants, ou bien rien de ce qui est enseigné n'est utile puisqu'on admet que n'importe quoi est suffisant. Dans les deux cas l'examen ne prouve rien, ce qui ne l'empêche pas d'exister.

La rédaction d'un tel catalogue d'exigences est un travail gigantesque, mais son utilité le justifie. Deux difficultés considérables attendent le(ou les) rédacteur(s) : ne rien oublier, jusqu'où aller ? Une assez bonne méthode pour oublier le moins de choses possible consiste à commencer par la fin : quelles activités devront-elles être requises pour que le diplômé n'ait pas une valeur exprimée en fausse monnaie ? Si le but des études est de savoir dépanner un ascenseur, le candidat dépanneur doit prouver qu'il sait dépanner dans la plupart des cas habituels.

Si l'un des buts des études est de "savoir utiliser la radio pour établir, confirmer ou éliminer un diagnostic", on peut ainsi le décomposer : quels sont les comportements constituant l'activité terminale souhaitée ?

. La radio a dû être prescrite dans les conditions appropriées(incidences) : SYNTHESE(élaboration d'un plan d'action).

. Pour choisir les conditions appropriées, il faut les connaître : I 25 connaissance des techniques.

. La recherche des signes constitue une analyse typique : recherche des éléments, recherche des relations entre éléments, recherche des principes d'organisation.

Les "signes" recueillis par la lecture

. peuvent être connus(s'ils sont "typiques") ; ils sont alors rapportés(II 20 interprétation) à des connaissances de structures concrètes (I 23)

. n'avoir jamais été vus : la référence peut alors être l'anatomie normale(I 23) et la prétendue interprétation est alors une application(III)

. peuvent n'avoir jamais été vus exactement de cette manière : ils seront alors rapprochés d'une structure déjà connue(I 23), ce rapprochement constituant une transposition(II 10)

. Avant d'être utilisés, les signes doivent être reconnus comme fiables : critique(VI) qui constitue une application de critères devant être connus(I 24)

Un certain tableau suggéré par l'analyse peut s'avérer incomplet, faute d'une incidence seule capable de révéler la présence(ou l'absence) du signe qui prend, ici, toute son importance : il faut donc connaître (I.25) l'incidence à utiliser et savoir la choisir dans un cas concret reputé nouveau : c'est une application(III). Le jugement final(diagnostic) et son expression(compte rendu) sont des synthèses(V) qui ne sont réalisables que par des sujets possédant certaines connaissances parmi lesquelles on peut souligner celles du vocabulaire(I.11) et des faits élémentaires(les signes) (I.12).

Pour s'en tenir à ce qui vient d'être décrit et qui est quelque peu simplifié, on peut retenir la plupart des objectifs élémentaires comme utiles et même nécessaires. S'ils sont écrits, il y a peu de risques que soient oubliées des choses aussi élémentaires que la connaissance du vocabulaire.

Quant à la question de savoir jusqu'où il faut aller, elle se ramène en fait, à une limitation quantitative au sujet de laquelle l'opinion du spécialiste est à priori suspecte. Si le spécialiste risque d'aller trop loin, l'utilisateur risque de ne pas aller assez loin, suspect qu'il est d'être enfermé dans une routine restrictive. Une assez bonne méthode pour savoir jusqu'où il faut aller consisterait à confronter plusieurs avis : il est difficile à un rédacteur de voir "juste" s'il ne s'est pas informé dans des conditions telles que les informations mêmes ne soient pas suspectes, mais à cet égard puisqu'on sait(assez bien) s'informer dans des domaines qui ne sont pas toujours très simples, pourquoi ne saurait-on pas aussi bien dans le domaine de l'enseignement ?

L'enquête aux fins d'information pourrait être fondée sur une base taxinomique. Ses résultats restent encore du domaine de la conjecture, car si elle a été faite, elle conserve encore un caractère confidentiel. On peut cependant penser que si les réponses sont accompagnées d'une identification des répondeurs, un certain nombre de choses cohérentes peuvent se constituer et dans la mesure où elles peuvent se montrer cohérentes, il serait sage d'en tenir compte. L'atomisation des objectifs qui résulterait d'une élaboration taxinomique entraînerait de nombreuses répétitions qui ne seraient pas sans intérêt. Il serait assez simple de regrouper certains objectifs en thèmes et certains thèmes pourraient aisément se regrouper en discipline ; la méthode d'élaboration ne préjuge pas de la forme du produit fini : la taxinomie est une méthode très sûre de choix des objectifs, ce n'est pas nécessairement la méthode la plus claire de présentation du classement. Il n'est pas nécessaire de connaître la recette pour apprécier un bon plat.

Un certain nombre d'objectifs peuvent être communs à plusieurs thèmes, et surtout, plusieurs thèmes à plusieurs disciplines. D'autre part, le niveau exigé n'est pas forcément le même dans toutes les disciplines.

Cette constatation probable entraînerait une conséquence heureuse : la découverte que plusieurs enseignements sont faits plusieurs fois, sans qu'il soit démontré quela répétition soit un avantage ni qu'une exigence identique s'exprime en épreuves comparables, faute d'avoir été définie. Si plusieurs thèmes absolument identiques sont revendiqués par plusieurs disciplines, il semblerait sage que chacune ne l'enseigne pas. Si la répétition est considérée comme un mode d'action utile, il faut discerner une différence entre "répéter" et"rabâcher" et savoir choisir entre répéter quatre fois et libérer du temps pour faire autre chose et ceci, d'autant que cinq cours sur le même sujet ne constituent pas nécessairement une répétition s'ils ne sont pas conçus de la même façon. Il est donc possible de concevoir "un programme" divisé en termes de disciplines et pour chacune d'elles, subdivisé en thèmes, chacun d'eux étant "atomisé" en objectifs élémentaires.

Un choix doit être fait, peut-être arbitraire, de la place d'un thème donné dans une discipline ou dans une autre. Si le niveau exigé n'est pas le même pour un thème donné dans plusieurs disciplines et sous réserve de la justification du niveau d'exigence, on a le choix entre une répétition ou une fragmentation du thème en objectifs plus ou moins ambitieux. Quel que soit le choix retenu, l'organisation se fait en fonction du temps : c'est un programme d'enseignement.

PROGRAMME D'ENSEIGNEMENT

C'est ici, dans son sens moderne qu'il faut prendre le mot. Alors que le programme d'objectifs est la liste des buts à atteindre, un programme d'enseignement est à la fois la liste des moyens à utiliser et l'organisation dans le temps des buts et des moyens appropriés pour atteindre au mieux ces buts : c'est une méthode. Le choix des moyens doit se fonder non seulement sur la nature du but, mais encore sur la qualité de la population enseignée. Dans l'hypothèse où un enseignement de littérature devrait être organisé, il ne suffirait pas que soit connue la liste de tous les comportements exigibles : la conception même du curriculum ne pourrait être la même si les étudiants sont licenciés ou analphabètes et le temps nécessaire à la réalisation du même programme risque d'être sensiblement différent. C'est dans la seule perspective des rapports avec l'examen que sera envisagé l'enseignement.

On sait que trois organisations temporelles sont possibles et le choix de l'un de ces schémas implique une certaine conception des rapports entre les disciplines qui explique leur distribution dans le temps.

1 - ENSEIGNEMENT TRADITIONNEL

2 - ENSEIGNEMENT PAR BLOCS

3 - ENSEIGNEMENT INTEGRE

1 - ENSEIGNEMENT TRADITIONNEL

L'enseignement dit traditionnel, et qu'il vaudrait mieux qualifier d'habituel, consiste à distribuer les activités tout au long de l'année, par petites tranches horaires équitablement réparties entre les différentes disciplines. C'est assurément une méthode administrativement commode, puisqu'à un moment donné, aucun groupe d'enseignants ne peut être déclaré favorisé par rapport à un autre et c'est un système d'une admirable simplicité pour répartir le temps à priori défini : c'est une excellente façon de constituer un emploi du temps, mais ce n'est pas réellement une méthode.

Il n'est pas du tout certain que l'exercice qui consiste à être anatomiste le lundi, physiologiste le mardi et chimiste le mercredi, soit à la portée de tout le monde, ni encore mathématicien le matin et historien l'après-midi. Il n'est pas plus certain que l'aptitude à un tel exercice représente un type de structuration souhaitable. Il semble évident que, dans de telles conditions, l'examen exerce une fonction assez particulière, dont il n'est pas démontré qu'elle soit souhaitable.

Ou bien l'enseignement a été dispensé toute l'année et l'année se terminant à la même date pour tout le monde, au même moment devront être remémorées (si le contrôle des connaissances est jugé suffisant) des faits, du vocabulaire, des structures, des lois, des théories qui risquent de n'avoir entre eux que de lointains rapports. Si on accorde à l'épreuve une fonction de stimulation, toutes les stimulations étant portées en même temps, on peut penser que les étudiants sont réveillés, au cas où l'enseignement n'y aurait pas suffi, mais il n'est pas inconcevable qu'ils soient épuisés et que, finalement, l'épreuve vérifie surtout la résistance. Cette qualité n'est pas à dédaigner mais il faut savoir ce que l'examen est censé vérifier réellement.

On peut encore penser que le défaut d'un tel système provient de l'éloignement très divers des enseignements par rapport à l'examen.
On pourrait estimer que l'épreuve centrée sur la partie terminale est plus facile puisque les acquisitions sont fraîches, donc encore présentes à la mémoire. On peut tout aussi bien estimer favoriser les candiats en leur proposant une épreuve centrée sur la partie initiale qui était plus simple et qui a plus de chances d'avoir été "mûrie"(comprise) le temps aidant.
On peut alors penser qu'il n'est pas logique d'exiger une compréhension plus rapide d'un enseignement plus complexe, ni de déterminer l'importance d'une question par la date de son enseignement.

Il est encore possible de pallier, à cette sorte d'inconvénients en multipliant les épreuves de telle sorte que chacune soit convenablement éloignée dans le temps de l'enseignement dont elles ont pour fonction de contrôler l'acquisition. Dans le système considéré d'enseignement, il n'est pas du tout certain que l'examen ne joue un rôle plutôt néfaste, car il semble évident que, plus il est proche, plus sera grande la proportion d'activité dépensée à sa "préparation" ... au détriment d'une discipline dont un examen est déjà passé ou encore lointain. Au total l'année se passe de bachotages en bachotages, favorisés par l'effet de "masque" de l'examen. Que cela soit ou non officiel, la veille(qui peut durer plusieurs jours) d'une épreuve, toutes les disciplines non concernées sont en vacances : c'est le calendrier des examens qui tient lieu de programme d'enseignement.

X

X X

2 - BLOC TEACHING

L'enseignement par "blocs"(le bloc - teaching des anglo-saxons) apporte un appréciable progrès au moins par l'effet de coordination et d'organisation qu'il exige. Il s'agit de consacrer la totalité du temps à une seule activité dans un intervalle donné. Cela oblige à construire un "ensemble" de leçons ou d'autres activités, de façon telle que, à cet échelon au moins, il soit possible de parler de "curriculum". Il est un peu plus difficile que dans le type précédent, de faire des exposés totalement étrangers les uns aux autres et cette contrainte appliquée aux enseignants n'est pas en soi, une mauvaise chose.

Dans ces conditions, examen et enseignement peuvent être coordonnés et différentes sortes de combinaisons sont possibles. L'inconvénient qu'on oppose en général à ce système, c'est qu'il engendre une certaine fatigue chez les étudiants consécutivement à la monotonie. Cet argument ne semble pas décisif. En effet :

a) Un "bloc" n'est pas nécessairement défini par une discipline. La définition d'une discipline par un contenu(dit"programme") est en soi, suffisamment arbitraire, pour qu'il soit acceptable de le diviser en thèmes. Il n'est pas indispensable de concevoir un "bloc de physiologie", on peut imaginer des blocs : la physiologie endocrinienne ici, la neurophysiologie là et la digestion ailleurs.

b) La fragmentation en blocs multiples d'une seule discipline ne la dénature pas. Dans un bloc unique, il ne sera pas possible de traiter en même temps de deux sujets dont les rapports sont pourtant évidents, telles la circulation et la respiration, sauf si on dispose d'une installation stéréophonique. Il est rare que ce genre de "synthèse" soit fait, mais rien n'empêche qu'on ajoute à deux blocs élémentaires un troisième, résultant de la réunion des deux autres dans ce qu'ils présentent d'utile à réunir.

c) La plus grande fatigue des étudiants peut être évitée par la nummulisation des disciplines dans la mesure où c'est la constance de l'objet d'application qui en est la cause et non la constance de l'effort. A ce sujet, cette méthode est plus exigeante que la précédente, ce qui ne serait un mal que si l'exigence était superflue.

Le choix du "bloc-teaching" expose à quelques inconvénients réels. Il serait vain de cacher que l'un d'eux consiste à considérer que, dans un intervalle de temps donné, la plupart des enseignants se trouvent "en vacances". Il semble scandaleux que cette catégorie de personnel, bénéficiant déjà de nombreux congés, passe une grande partie de son temps hors congé à ne rien faire. Cette vue des choses mérite considération : elle signifie que la seule activité des enseignants est d'enseigner.

Ce point est déjà discutable à un certain niveau d'activité : on peut soutenir ou réfuter qu'un enseignant soit encore chercheur ou exerce une autre activité en rapport avec sa discipline. En admettant pour simplifier, que l'activité d'enseignement soit exclusive, il resterait encore à savoir ce que cela signifie. Si on ne mesure l'activité d'enseignement que par le nombre d'heures passées en présence d'un public étudiant, on fait une grave erreur. En poussant cette conception aux extrémités permises par la logique, un professeur pourrait passer quelques heures dans une cabine d'enregistrement et dire ses leçons qu'il ne resterait plus qu'à diffuser aux heures fixées par le "programme". Accepter ce point de vue reviendrait à supprimer toute préparation, toute évolution, toute adaptation à une population donnée et tout travail d'élaboration et de "correction" des examens, ce qui reviendrait à dire que toutes ces activités sont inutiles. En réalité, il est préférable de s'en tenir aux réalités plutôt qu'aux apparences et d'éviter que les secondes ne remplacent les premières.

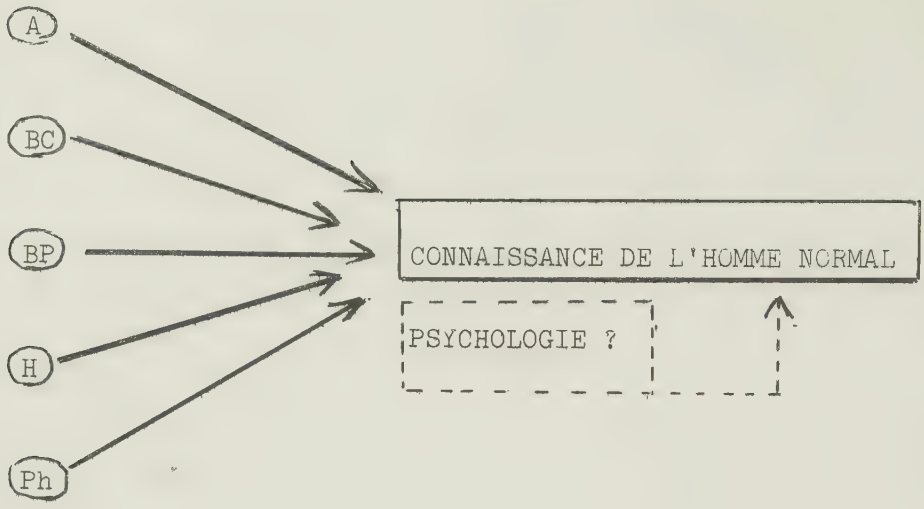
Un autre inconvénient, beaucoup plus grave, puisqu'en rapport avec l'enseignement même, s'attache au "block-teaching" pris dans son sens le plus rétréci : celui qui consiste à juxtaposer des disciplines sans aucun souci d'en faire un tout.

Soit, dans une année par exemple, à enseigner cinq disciplines : l'anatomie, la biochimie, la biophysique, l'histologie et la physiologie. En admettant que l'année universitaire dure quarante semaines, il est possible de bloquer chacune des disciplines sur huit semaines et par exemple, de les répartir par ordre alphabétique. La séquence réalisée sera :

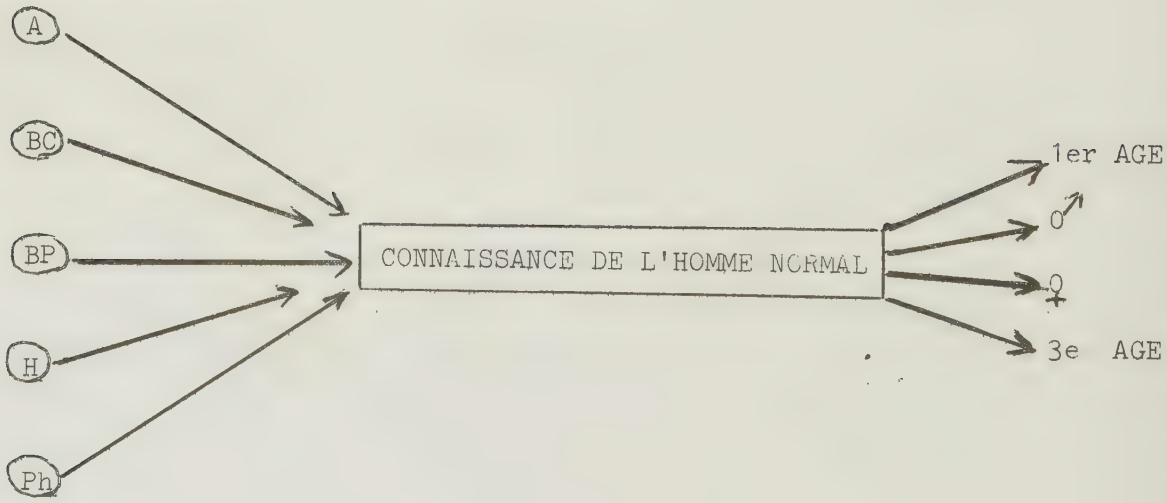
$$A \longrightarrow BC \longrightarrow BP \longrightarrow H \longrightarrow Ph$$

et leur répartition dans le temps implique :

1) que les difficultés d'apprentissage sont les mêmes puisque les temps impartis sont égaux.


A
BC
BP
H
Ph
CONNAISSANCE DE L'HOMME NORMAL
PSYCHOLOGIE ?

N° 1.


A
BC
BP
H
Ph
CONNAISSANCE DE L'HOMME NORMAL
1er AGE
♂
♀
3e AGE

N° 2.

2) ou bien que les difficultés sont décroissantes puisque le temps passant, les étudiants sont de plus en plus fatigués.

3) ou bien que les difficultés sont croissantes puisque le temps passant, les étudiants sont de plus en plus entraînés.

4) que les examens doivent être répartis dans l'ordre des enseignements, et non groupés à la fin de l'année, sans quoi le temps théorique prévu pour la dernière discipline serait en fait, exclusivement consacré aux préparations d'épreuves.

5) que chacune de ces disciplines est indépendante des autres.

En soulignant le défaut de la division systématique en temps égaux, on peut encore soutenir qu'il n'est pas possible de résumer l'opération par la justification émise à posteriori, ou même a priori, et consistant à dire que les cinq disciplines ont pour objectif commun la "connaissance de l'homme normal" par ce qu'il n'est pas du tout certain que :

1) chacun des professeurs ait construit son cours dans cette perspective.

2) les étudiants aient eux-mêmes découvert les rapports utiles entre l'anatomie(oubliée, examen passé) et la physiologie(ultime obstacle à franchir).

Si le but réel était "la connaissance de l'homme normal", le schéma de répartition précédent serait remplacé par le schéma d'organisation n° 1.

La réalisation d'un tel schéma nécessite une certaine coordination et impose un programme. Il devient évident que les cinq disciplines imposent une certaine conception, restrictive, de la description et de la compréhension et la question se pose de l'utilité d'autres disciplines, par exemple la psychologie. Enfin, la description du but semble un peu trop concise : les moindres choses qu'on ait envie d'ajouter sont en rapport avec "homme normal". Si on accepte l'expression dans le sens de "être humain normal", de nouveaux buts apparaissent : il est normal d'être enfant, adulte ou vieillard, homme ou femme. Le but se fragmente en âge(1er, 2ème, 3ème) et en sexe et le schéma devient tel qu'en N° 2

Un travail d'organisation s'impose, d'autant qu'il est nécessaire de définir ce qu'on entend ici, par "connaissance".

Soit à concevoir un enseignement par "blocs" à un niveau relativement élevé(2ème partie du second cycle) des études médicales. Il s'agira, en résumé, d'atteindre à la "connaissance de l'homme malade". Cette opération peut se décrire. La connaissance de l'homme malade c'est la pathologie qui décrit des "tableaux" particuliers qui sont des maladies identifiées; "Pathologie humaine" possède un sens restrictif adéquat au genre d'études considérées. Connaître la description de ces tableaux est une acquisition évidemment nécessaire pour le futur médecin, mais cette connaissance ne résume pas la fonction du médecin : il doit découvrir les signes présentés par son malade, les regrouper en un tout à comparer avec un tableau mis en mémoire (diagnostic) et traiter le malade. Trois disciplines sont mises en jeu : connaître, savoir rechercher et comprendre les signes(séméiologie), connaître les maladies(pathologie), savoir les traiter dans la mesure du possible actuel(thérapeutique).

Chacune de ces disciplines est très naturellement, traditionnellement et logiquement, subdivisée en thèmes, tels la cardiologie, la néphrologie, la pneumologie... etc. Soit trois "thèmes" a, b, et...n et S, la séméiologie, P, la pathologie, et T, la thérapeutique. Deux conceptions de l'enseignement par "blocs" peuvent s'imaginer :

A) DES GROS BLOCS où la discipline est l'unité :

1) $Sa \rightarrow Sb \longrightarrow ..Sn = S$

2) $Pa \rightarrow Pb \longrightarrow ..Pn = P$

3) $Ta \rightarrow Tb \longrightarrow ..Tn = T$

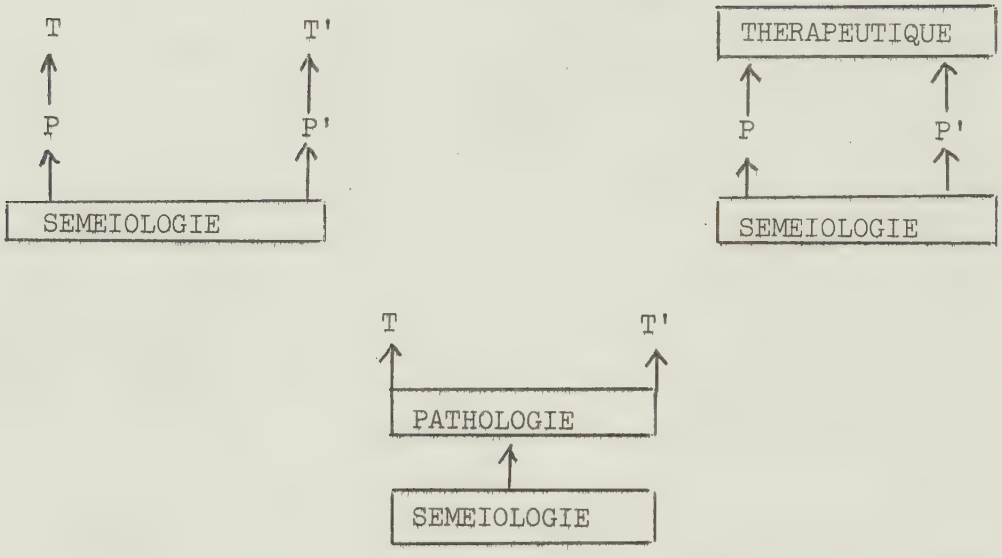
Ainsi sont enseignés successivement des signes isolés(S) des signes groupés(P) et le traitement des maladies(T). Cet ordre présente une certaine logique, mais peut-être, quelques inconvénients :

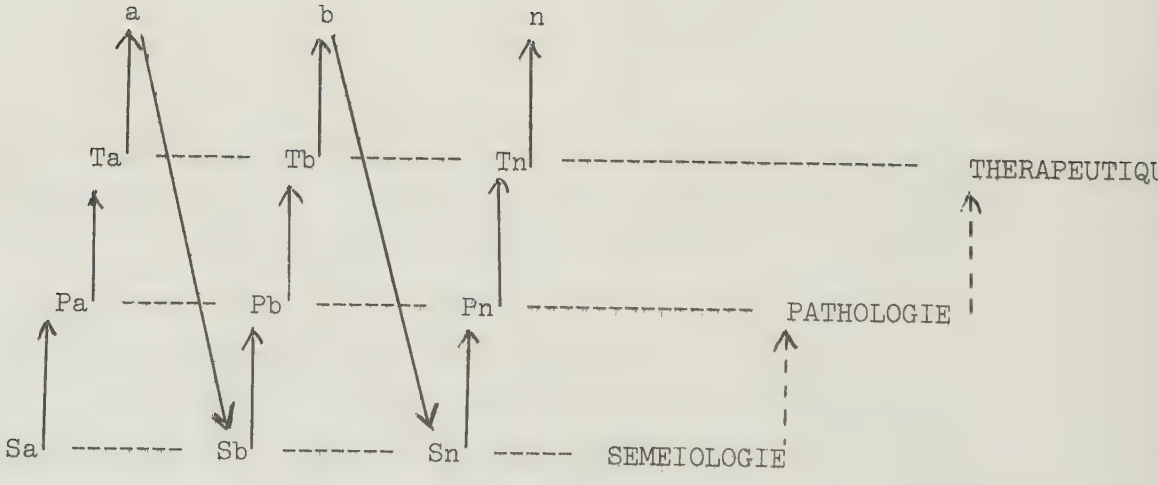
a) chaque bloc est lourd.

b) le but reste longtemps lointain puisque on a l'impression de l'atteindre, de savoir réellement quelque chose, lorsqu'on arrive à la thérapeutique, geste significatif qui pourrait bien représenter la réelle motivation des études entreprises.

c) Il n'est pas du tout certain que l'importance d'un signe soit réellement acceptée tant qu'il n'est pas placé dans sa perspective réelle : le moyen de parvenir à un diagnostic précis, exercice interdit jusqu'à ce que soit abordé le bloc de pathologie.

QUELQUES VARIANTES où l'unité est variable.





———— Schéma du bloc B (petits blocs)

---------- Schéma du bloc A (gros blocs)

N° 3.

Si un tel choix peut se défendre, avant de le faire il est bon de se demander pourquoi réellement on le fait. N'est-ce pas dans le but de "ménager des effets", dans la perspective d'une "progression dramatique", avec l'idée inavouée de "retarder" ce qui intéresse réellement, pour tenir en haleine... pour retarder, simplement ? Il ne semble pas que de tels mobiles soient les meilleurs possibles ni qu'on puisse tout à fait en écarter l'hypothèse.

B) DES PETITS BLOCS où le thème est l'unité

1) Sa + Pa + Ta ⟶ a

2) Sb + Pb + Tb ⟶ b

.....

n) Sn + Pn + Tn ⟶ n

Ainsi sont enseignés successivement les thèmes a, b...n. La logique qui préside à cette distribution a l'incontestable avantage d'être en rapport avec le but défini des études et les inconvénients du système précédent disparaissent au moins en partie. Cependant, il n'est pas certain qu'un tel enseignement soit aussi bien calqué sur la réalité qu'on pourrait le croire : il risque d'en fournir une représentation un peu trop simplifiée, mais cet inconvénient n'est pas sans remède dans la mesure où il est admis que ce qui a été acquis au cours d'un bloc doit non seulement rester présent au moment de l'étude des suivants, mais encore être renforcé.
Rester présent et être renforcé représentent des choses sensiblement plus faciles à dire qu'à réaliser et à vérifier.

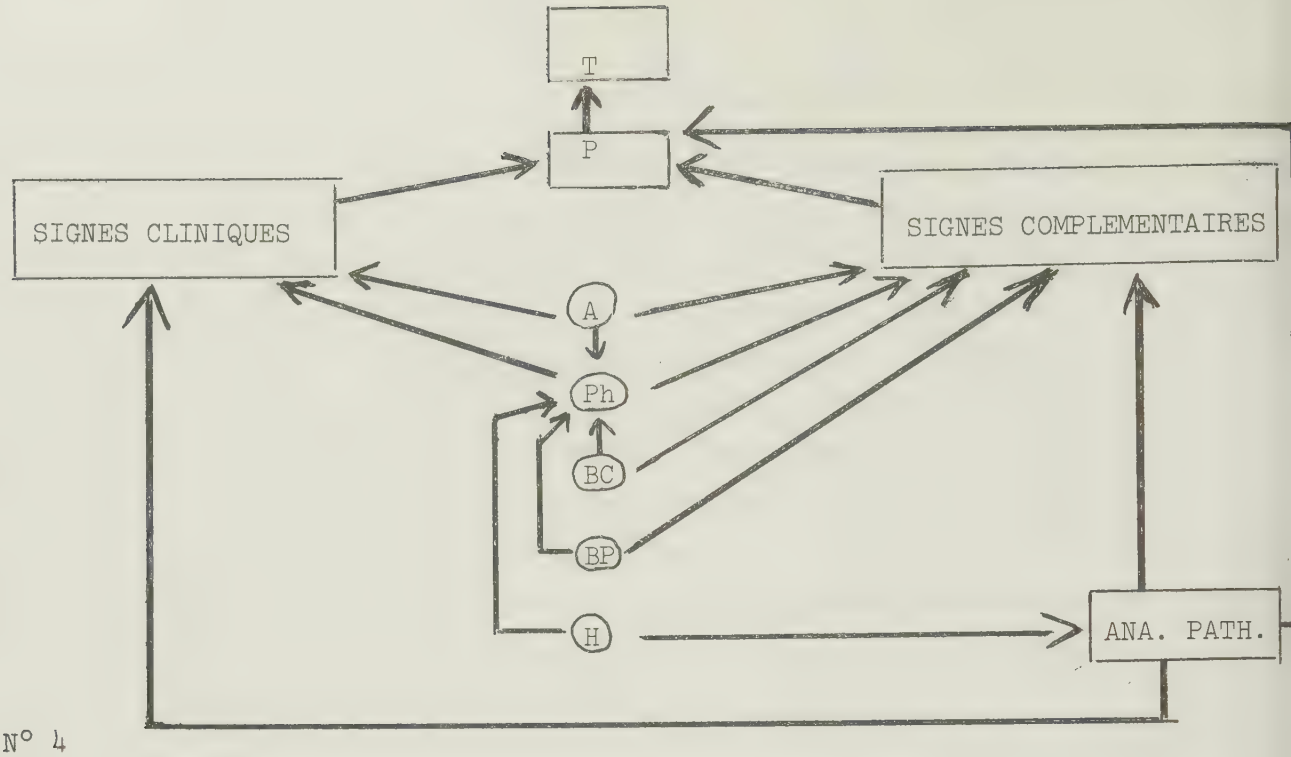
Au total, deux schémas fondamentaux de "bloc-teaching" peuvent être proposés. Le choix de ce système impose une coordination d'autant plus étroite que le bloc est choisi plus petit et donne aux enseignements une certaine unité. A chacun de ces schémas correspond un choix de programmes d'examens(Schéma n° 3)

3 - ENSEIGNEMENT INTEGRE

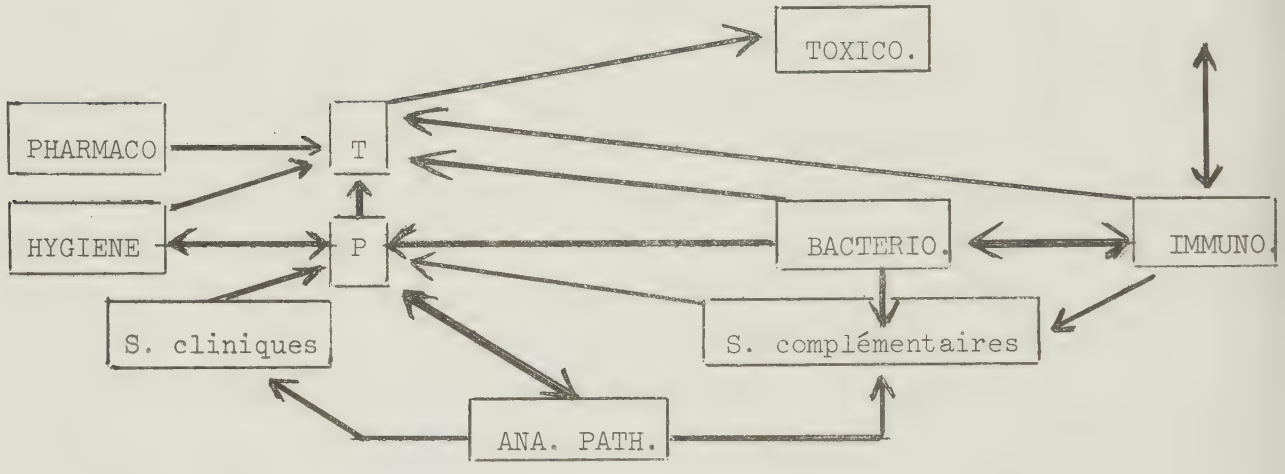
L'enseignement intégré représente l'organisation la plus complexe puisque l'unité n'en est plus la discipline, mais le thème : le nombre d'unité à distribuer est donc par définition, plus grand. Contrairement à ce qui est souvent pensé, le schéma B d'organisation en blocs est bien un schéma de bloc et non d'enseignement intégré. Alors que, dans le bloc , l'unité prise en considération est un thème, lui-même subdivision de discipline, dans l'enseignement intégré on procède par grands thèmes, eux-mêmes divisés en thèmes de moindre importance qu'on ne peut placer dans une discipline donnée que par la force de l'habitude : la notion de discipline traditionnelle éclate littéralement, ce qui n'est pas nécessairement un inconvénient, mais le seul fait d'éclater une structure traditionnelle n'est pas une garantie suffisante d'avantages.

Soit à concevoir l'enseignement intégré d'une pathologie, à dessein, indéterminée. Le but est de savoir la traiter(thérapeutique). Pour cela il faut savoir la reconnaître(pathologie) et pour la reconnaître, il faut savoir en rechercher les signes(séméiologie), les regrouper, les interpréter et les comprendre. Les signes à savoir rechercher sont de deux ordres : cliniques(quête sans instruments) et non cliniques ou complémentaires dont la recherche nécessite une instrumentation plus ou moins complexe, parfois fort complexe, et la collaboration d'un laboratoire qui se charge de la technique et qui fournit les résultats... qu'il ne reste plus qu'à comprendre. Jusqu'à "la compréhension", le schéma restait celui d'un bloc à ce détail près que la séméiologie a été scindée en clinique et complémentaire. C'est au-delà que le schéma change car il faudra recruter toutes les compétences nécessaires à la compréhension de certains faits dans une perspective précise.

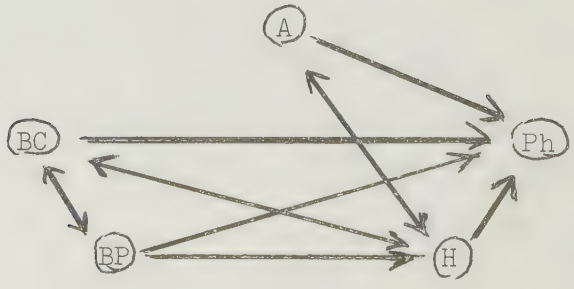
Ainsi, la séméiologie clinique pourra requérir des connaissances particulières en anatomie, en physiologie et en anatomie pathologique. L'anatomie pathologique permet de mieux comprendre la pathologie ; elle constitue un examen complémentaire et ne peut être comprise qu'à partir de connaissances en histologie. La séméiologie complémentaire peut requérir des connaissances précises en anatomie(radio), en physiologie(exploration fonctionnelle), en biochimie(beaucoup d'examens de laboratoire), en biophysique (certains examens de laboratoire tels ceux qui utilisent des radio - isotopes).



N° 4



N° 5
N.B. On peut remplacer " BACTERIOLOGIE " par VIROLOGIE, PARASITOLOGIE...



N° 6

Enfin, certaines connaissances en physiologie, par exemple, ne sont accessibles qu'à partir d'autres connaissances en anatomie, biochimie, biophysique, histologie.

L'organigramme de l'enseignement considéré peut se représenter comme en 4

Il n'est pas simple et pourtant, il n'est pas complet : ont été négligés, par exemple, les rapports entre la biochimie et la biophysique d'une part avec l'histologie d'autre part, et encore l'implication des disciplines "fondamentales"(entourées) dans la thérapeutique.

D'ailleurs, le passage de la pathologie à la thérapeutique est représenté de façon simpliste. Soit, par exemple, à concevoir l'organigramme de la pathologie infectieuse : une représentation, encore simplifiée, est proposée (5)

Apparaissent ici des rapports à double sens indiquant que certaines acquisitions d'une discipline "a" sont favorisées par des acquisitions préalables dans les disciplines "b" et inversement. Si on ajoute ce qui manque à l'évidence sur le graphique précédent : les disciplines fondamentales, on admet sans peine qu'une telle programmation est complexe d'autant que d'évidentes interrelations existent même au niveau des disciplines traditionnellement désignées comme "fondamentales"(cf. Schéma N° 6)

Au total, l'enseignement intégré est très complexe dans son organisation et pour mériter son nom qui suppose un haut degré d'organisation, il admet moins que tout autre, il n'admet pas l'improvisation. Il présente un immense avantage : il impose un sens à toutes les activités, ce qui suppose que le but en soit précisément défini. Ainsi, certains spécialistes ont deux fonctions : traiter certaines maladies et savoir reconnaître, grâce à une technique appropriée, certaines maladies qu'ils n'ont pas à traiter. Par exemple, l'ophtalmologiste est plus particulièrement compétent pour fournir certains renseignements sur l'hypertension, le diabète ou l'état de la circulation cérébrale. Peut-être que le recours à de tels spécialistes dans des disciplines qui ne sont pas les leurs, est la preuve d'insuffisances regrettables d'autres spécialistes, peut-être non : le choix reflète une certaine conception de l'exercice qui transparaît nécessairement dans une programmation logique.

Choisir ou refuser la conception de l'enseignement intégré est une chose, le réaliser en est une autre. On peut penser que son intérêt a des limites : on ne commence pas des études de médecine à la maternelle, et que son utilité est plus claire sur des étudiants ayant déjà une certaine culture et s'engageant dans un cycle d'études dont le but est clair. Non seulement ce type d'enseignement s'adresse plus à des étudiants qu'à des lycéens mais encore, parmi les étudiants, plutôt à ceux qui ont déjà atteint, par d'autres méthodes, un certain niveau, ce qui justifie la distinction de différents cycles et nécessite que ce qui est distingué soit défini. Enfin, même si jamais ne devait ou ne pouvait être appliquée une telle programmation, son élaboration serait, en soi, un grand progrès par les questions qu'elle soulèverait.

PROGRAMME D'EXAMENS

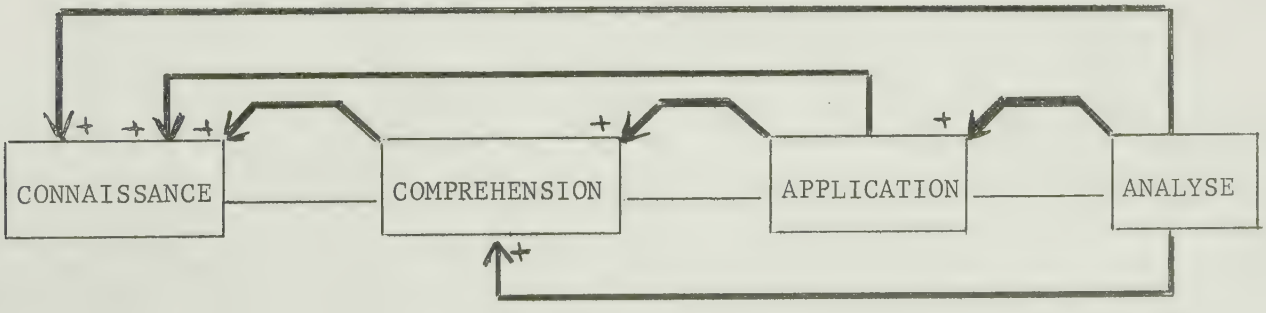
Contrairement aux objectifs d'examen qui doivent définir(se confondre avec) les objectifs d'enseignement, le programme d'examen doit être établi en fonction du programme d'enseignement.

Un tel programme n'est pas un calendrier, mais il est évident que son élaboration finit par aboutir à un calendrier. En dehors du choix qui Peut se faire en fonction du type d'enseignement retenu, et dont dépend le calendrier, un schéma général de conception peut être élaboré indépendamment de la nature de l'unité retenue : discipline ou thème.

Si le thème est ici retenu comme sujet de réflexion, le passage à la discipline ou à un ensemble de thèmes ou à un ensemble de disciplines rend les choses plus complexes par la masse plus grande de ce qui doit être vérifié. Ce sont les catégories taxinomiques qui serviront de guide puisque ce sont elles qui ont fourni la liste des objectifs et que ceux-ci sont définis en termes de comportement observables, vérifiables donc en termes d'examen.

Soit l'objectif proposé : "connaître telle théorie" : il est clair, le mot "connaître" ayant été choisi à dessein pour résumer le comportement souhaité, que l'épreuve est un "contrôle des connaissances" ou une "vérification des connaissances". Pour les étudiants, il s'agit d'arriver à la connaissance et pour les enseignants, de faire en sorte que la plupart des étudiants puissent arriver à la connaissance, ce pourquoi ils ont une grande liberté de moyens.

Une théorie a plus de chances d'être retenue(c'est cela la "connaissance") si elle est comprise : ainsi, la compréhension peut être inscrite au programme d'enseignement. L'application aide à la compréhension(qui aide à la connaissance) et l'analyse aide à l'application, à la compréhension (recherche des relations entre éléments et des principes d'organisation) et à la connaissance(recherche des éléments). Par ailleurs, il faut un peu de connaissance pour comprendre, de la compréhension pour appliquer des connaissances et de la compréhension pour analyser. Le schéma de l'organisation du cours peut se représenter ainsi :



L'analyse de l'enseignement le décompose en quatre éléments pouvant contracter diverses relations, mais le but à atteindre n'est que la connaissance. Si le professeur a jugé bon, utile, de dépasser le but proposé, l'épreuve elle, n'a pas à vérifier que l'enseignement a été suivi(après tout, le plan choisi n'est pas nécessairement le plus efficace), mais elle doit vérifier que l'objectif a été atteint indépendamment des moyens utilisés. Bien que le cours ait été analytique et peut-être même, synthétique, l'épreuve restera d'un niveau élémentaire : vérification des connaissances.

Si le but proposé est une critique, cela signifie que l'épreuve doit nécessairement vérifier la capacité des candidats dans le domaine de la critique concernant le thème choisi. Pour critiquer, il est nécessaire de savoir analyser ; aussi, est-il légitime qu'une épreuve de critique comporte une partie d'analyse. Légitime, mais non obligatoire.
En effet, s'il est demandé à un candidat de montrer qu'il est capable de critiquer, c'est sur cette seule capacité qu'il doit être jugé : il est capable ou non et une très brillante analyse ne pourrait que confirmer son incapacité de critiquer et on ne voit pas en quoi, logiquement, une incapacité patente pourrait lui permettre d'être accepté. Si, en revanche, l'épreuve n'est pas terminale, il peut être utile de vérifier la capacité à l'analyse dans le domaine choisi, puisque c'est une étape obligatoire. Un étudiant faisant une bonne analyse mais une mauvaise critique, serait différencié de celui qui ferait une mauvaise critique et une mauvaise analyse et, à ces deux étudiants, des conseils différents pourraient être donnés. Quant à la bonne critique précédée d'une mauvaise analyse, elle poserait la question de la fiabilité de l'épreuve, tant apparaîtrait suspecte et de peu de valeur une bonne réponse à un problème "obtenue" à la suite d'une mauvaise solution.

La place des examens, l'organisation des épreuves et leur progression, leur nombre sont autant de faits élémentaires dont la mise en ordre constitue un programme d'examens dont le choix dépend, pour une large part, du programme des études et de l'idée qu'on se fait de termes tels que "contrôle continu", "examen partiel", "examen terminal", "épreuve partielle", "épreuve terminale" et qui seront analysés plus loin.

PUBLICATION DES OBJECTIFS

Les buts doivent être définis et les buts définis sont des objectifs. Ces objectifs sont communs aux enseignants et aux étudiants. Il est évident que les enseignants doivent les connaître : dans la mesure où ils sont "pédagogues", il ne leur est guère possible de "guider" ceux qu'ils doivent guider s'ils ne savent pas vers quoi et la précision de leur action dépend, notamment, de la précision dans la définition du but.

Les étudiants doivent aussi en avoir connaissance et pour plusieurs raisons :

1°) le but doit être atteint par eux-mêmes et il semble de la plus élémentaire honnêteté de dire à des gens qui vont passer beaucoup de temps et dépenser beaucoup d'énergie "pourquoi" ils ont à le faire et de leur annoncer ce que l'on attend d'eux.

2°) surtout dans le cadre universitaire, le "travail personnel" occupe une place importante. Cela signifie que les étudiants ont une large liberté de manoeuvre et de choix des moyens et les meilleurs moyens possibles, variables en fonction de la personnalité du sujet, sont ceux qui sont le mieux adaptés aux fins : encore faut-il les connaître.

3°) qu'on le veuille ou non, les étudiants ne sont pas désintéressés : ils font des études dans un but précis qui est celui "d'avoir l'examen" même s'il n'y a aucun rapport entre la nature des épreuves et le but avoué des études(divergence dont la faute incombe aux enseignants).

Par des renseignements issus de la promotion précédente, par exemple, ils finiront par connaître ce que l'on attend d'eux et plutôt mal, par des informations de seconde main, dont la quête aura perdu un temps qu'il aurait été très facile d'épargner.

Lorsqu'on demande à des enseignants s'ils pensent que l'angoisse pertube le résultat de l'examen, ils l'admettent volontiers. Or, dans maints domaines, l'angoisse c'est la peur de l'inconnu et il ne semble pas qu'il en soit autrement pour un examen. Son enjeu est considérable puisque,au mieux, un échec coûte du temps et, parfois, une carrière.

Si, à l'importance de la partie, on ajoute systématiquement l'ignorance des règles du jeu, il est facile de comprendre que les conditions de sérénité requises pour un exercice intellectuel ne soient pas remplies. Il arrive que, jusqu'au dernier moment, des candidats ne sachent pas si leur épreuve sera écrite, orale ou pratique, s'ils auront à répondre à des QCM, à rédiger de longues ou de courtes réponses, s'il leur faudra répéter(réciter) leur cours, faire des exercices d'application, montrer s'ils ont compris et comment, si telle épreuve est prise en considération ou non, pour combien, si une erreur dans un certain genre de question est rédhibitoire ou non, s'il vaut mieux ne rien dire que de risquer l'erreur ou au contraire, si l'omission sera plus durement pénalisée que l'erreur, etc...

Un candidat à une épreuve de médecine avait à traiter d'une certaine manoeuvre à faire dans un cas précis et il l'a décrite correctement. Dans cette question, intitulée "conduite à tenir"... il avait oublié de nommer la manoeuvre et le résultat fut zéro. La conclusion c'est qu'il aurait mieux valu répondre par un mot que par une description contrairement à ce que laissait supposer la nature de la question . Il est compréhensible qu'une telle incertitude soit cause d'angoisse, admis que l'angoisse pertube les résultats et clair que de tels résultats faussés le soient faute de précision dans la proclamation de ce qu'on attend des candidats.

Enfin, la publication des objectifs présente encore un avantage considérable : c'est une sorte de contrat qui établit clairement ce que les examinateurs attendent des candidats et ce que les étudiants doivent attendre des enseignants. Si l'enseignant est l'examinateur, une telle publication lui interdit d'improviser en cinq minutes des questions portant sur un cours déjà lointain et qui peuvent n'avoir qu'un rapport incertain avec l'enseignement préalablement fait et qui constitue, pour les candidats, la seule source d'information. Si l'examinateur est indépendant de l'enseignant, il lui est interdit, de fait, d'improviser une épreuve résultant d'une interprétation tout à fait différente de celle qu'en a faite l'enseignant de ce qu'on appelle, à tort, "programme" et qui est, de plus, "officiel".

CHAPITRE X

CIRCONSTANCES D'UN EXAMEN

D - ETABLIR UN PROGRAMME

On peut considérer que les lauriers universitaires ne fleurissent qu'en été: c'est l'habitude; ou que l'université est une serre dans laquelle on peut obtenir une floraison à n'importe quelle saison : c'est nouveau; ou encore qu'il n'y a que de lointains rapports entre l'enseignement et la botanique, entre la culture et l'agriculture: c'est audacieux. Ces conceptions différentes se reflètent dans les interminables querelles à propos du " contrôle continu (des connaissances)" des épreuves ou examens partiels, des épreuves ou examens terminaux, de l'examen " global " ou des épreuves indépendantes.

CONTROLE CONTINU 331

EPREUVES PARTIELLES 334

EXAMEN TERMINAL 338

DES CIRCONSTANCES PARTICULIERES 341

Numerus clausus (341); Sélection (342); Sélection et Numerus clausus (343); Examen d'un effectif sélectionné (343).

ENSEIGNEMENT ET EXAMEN 345

UN PLAN D'EXAMEN 349

LE CONTROLE CONTINU

Ce terme, en France, fleure son révolutionnaire, ce qui enthousiasme les uns et repousse les autres mais ce n'est pas sur ce critère qu'il doit être jugé: si on le choisit, c'est qu'il présente des avantages; si on le refuse, c'est qu'il présente des inconvénients; par exemple, de ne servir à rien et de faire perdre du temps à tout le monde.

Inscrit dans la loi, ce terme n'y a malheureusement pas été défini ce qui, une fois de plus, permet toutes sortes d'interprétations. Pour les uns "continu" s'oppose à terminal, ce qui permet de déclarer sans rire qu'un contrôle continu est constitué de deux épreuves partielles subies en cours d'enseignement dans un cycle de seize semaines: c'est une notion de la continuité qui semble relativement fausse. Parfois, on peut assister à de très sérieuses discussions (très longues en tous cas) sur le point de savoir si deux épreuves suffisent pour parler de "contrôle continu" et s'il ne serait pas plus convenable d'en proposer trois!

Pour les autres, c'est un grand nombre d'épreuves qu'il convient d'organiser pour garantir la continuité, à la limite, une fois par jour: cette conception mérite considération. Pour d'autres, enfin, le contrôle continu s'oppose à l'épreuve terminale par sa forme. Un groupe d'étudiants est confié à un enseignant (assistant) parfois, mais pas toujours, au même dans une discipline donnée ou l'une de ses subdivisions. L'assistant interroge ses étudiants oralement ou par écrit, et les note. Cela peut être une excellente ou une détestable conception selon l'usage qu'on fait des notes.

Ou bien les notes attribuées sont indicatives et la méthode est excellente, ou bien les notes "comptent" pour une fraction souvent importante (40%) dans le résultat final et le système est détestable. Dans le meilleur des cas, un Professeur ayant le souci de coordonner l'enseignement dans les différents groupes de l'effectif dont il est responsable, ayant une claire conscience de la gravité de la sanction d'un examen (positive ou négative, même positive) ayant défini ses objectifs à des collaborateurs qu'il a choisi et en qui il a confiance, ne peut pas prendre en considération,pour les faire entrer dans des additions, des notes dont il ne peut savoir ce qu'elles représentent.

Ou bien les interrogations ont été écrites et la lecture des copies ne peut suffire si on ne sait pas comment s'est déroulée l'épreuve improvisée. Il reste encore à la noter et à comparer les résultats des différents groupes qui n'ont pas eu à traîter des mêmes questions mais qui concourent cependant au même but: au même examen. Ou bien les interrogations ont été orales et le seul document restant est une impression notée par l'interrogateur. L'expérience montre que, même en accordant une totale confiance aux assistants, en admettant qu'ils ne jugent pas de la même façon une situation identique, la pondération de leurs notes est injuste. Un même assistant peut avoir à connaître plusieurs groupes qui se montrent d'inégales valeurs et, surtout, un même groupe peut être jugé différemment par plusieurs assistants pour la simple raison qu'une personne, quelle que soit sa qualité, peut être acceptée par un groupe et refusée par un autre. Dans ces conditions, il est clair que la note, a une signification mais dans une certaine mesure, peut être dans une large mesure, pas celle qu'on croît. Hormis le cas, à peine imaginable, où le même enseignant aurait à faire à tout l'effectif réparti à chaque fois en groupes différents, de telles notes n'ont pas de signification suffisamment claire pour qu'on ose les additionner.

Cela ne signifie pas qu'une telle méthode de contrôle soit dépourvue de toute valeur. D'une part, il est certain que n'importe quelle forme de vérification peut apporter des informations. Pour les étudiants, ces informations aboutissent à des conseils; encore faut-il qu'ils osent montrer leurs insuffisances, qu'ils ne cherchent pas à tromper l'interrogateur sur leur valeur réelle, ce qui n'est concevable que si aucune sanction ne s'attache à l'opération. Si, au contraire, la note est "prise en compte", le jeu normal des étudiants est un jeu de candidats: obtenir le plus possible à partir d'une performance donnée, en soignant la forme et en flattant les goûts de l'interrogateur connu. Pour les enseignants, il est possible d'utiliser les informations recueillies aux fins de savoir sur quoi faire porter le travail après avoir dûment constaté où l'aide serait utile. Pour les uns comme pour les autres, le résultat qu'on peut espérer est une amélioration: une correction, et c'est dans le sens anglais qu'il convient de comprendre le mot "contrôle" . Un tel contrôle ne peut être efficace que s'il est continu.

Rien n'empêche l'assistant de tenir à jour ses impressions, avec le maximum de précision et celà peut servir. Inéluctablement, le jury devra arbitrairement juger un certain nombre de candidats que les notes seules ne peuvent déclarer acceptables.

On ne peut rêver meilleures informations pour un jury que celles qui viennent des enseignants. Ce n'est pas le jury qui peut savoir que tel candidat se révèle toujours très solide en épreuve orale mais peu brillant dans une épreuve écrite, ou que tel autre est régulièrement satisfaisant, contrairement à ce que laissent supposer ses notes d'examen ou telle note d'épreuve. Le jury n'est évidemment pas contraint par ces renseignements mais il peut être suffisamment alerté pour décider d'un complément d'information avant de prononcer une sanction. Jamais le jury ne pourra se plaindre d'avoir un dossier trop bien fourni pour l'aider à prendre position et le contrôle continu - même noté et avec toutes les imperfections qu'il suppose - peut-être, à cet égard, d'une aide précieuse et d'une utilité incontestable.

Au total, il faut choisir entre la vérification aboutissant à une sanction et le contrôle continu n'a pas sa place dans un examen (il n'est pas fiable) et perd sa place dans l'enseignement (il ne peut plus corriger). Une vérification, aboutissant à une sanction et qui ne serait pas "totale" serait partielle.

EPREUVES PARTIELLES

Il faut souligner, pour les éliminer, quelques subtilités de vocabulaire qui empoisonnent les discussions. La grande différence entre "contrôle continu" et "épreuves partielles" vient de ce que l'une des opérations est définie par une intention... mal définie(contrôle) l'autre par une forme (épreuve) presqu'aussi mal définie, mais beaucoup plus définissable. Il n'est pas du tout inconcevable que des épreuves partielles soient organisées dans un but de contrôle, aussi bien aux fins de vérification qu'aux fins de régulation : ces dernières ne pouvant pas constituer un objectif pour une épreuve terminale. Au nombre et au sujet près, il n'y a pas de différence dans la conception entre les épreuves, qu'elles soient terminales, ou initiales, ou intermédiaires, globales ou partielles. Des différences considérables apparaissent, en revanche, si on considère les impératifs de la notation et la nécessité d'une organisation évidemment plus exigeante pour un ensemble d'épreuves que pour une épreuve isolée. C'est pour souligner le caractère nécessaire de l'organisation que seront appelés :

- partiel, tout ce qui ne sera pas "complet" ,
- épreuve, une méthode définie d'obtention de résultats,
- contrôle, les informations fournies par les résultats et leur éventuelle utilisation.

Quelques avantages sont évidents en faveur des "partiels". Le premier, c'est de limiter la part du hasard. On a moins de chances de se tromper en effectuant dix fois une mesure qu'une seule fois ; à condition de ne pas changer d'unité d'une opération à l'autre. D'autre part, dans la mesure où on est exigeant pour chacune d'entre elles, chaque épreuve restreinte se combinant à une autre, les chances de réussite d'une impasse se trouvent notablement réduites. Encore ne faut-il pas transformer le jeu en une chasse systématique à l'impasse, quelle qu'elle soit : certaines sont raisonnables, d'autres non. Le discernement de l'examinateur n'est pas un facteur de protection suffisant contre les excès : des objectifs définis, oui.

Le second est de pouvoir mettre en évidence, de manière rapide sinon immédiate, une tranche d'enseignement "mal passée". Dans la mesure où "le trou" est le même pour tout le monde, les probabilités sont grandes pour que l'enseignement soit responsable : il est encore temps de rectifier. Cela suppose que quelqu'un sache de quoi doit être probante l'épreuve et, par conséquent, que ses objectifs soient définis !

Le troisième est de pouvoir se rendre compte d'une imperfection dans la forme de l'épreuve sur laquelle il est possible de revenir: une telle erreur est réparable et non dommageable. Il n'en serait pas du tout de même pour une épreuve unique au cours de laquelle des candidats pourraient apprendre et fixer des notions fausses : cet aspect très mal connu de l'influence des examens sur le savoir (doxologie) pourrait être précisé grâce à l'étude des résultats de nombreux partiels, chacun étant centré sur un nombre restreint d'objectifs définis.

Le quatrième est de pouvoir perfectionner les méthodes d'enseignement jugées sur les résultats qu'elles fournissent, en fonction d'objectifs définis. Une épreuve terminale ne peut pas fournir des informations aussi sûres. Se fier aux seuls examens terminaux pour juger des résultats d'un enseignement constitue une perte d'information puisque la population suivante sera différente. Celà reviendrait à régler le chauffage à l'aide d'un thermostat qui tiendrait compte de la température moyenne de l'année précédente.

Le cinquième est de supprimer le caractère solennel de l'épreuve. Elle perd certainement de son caractère initiatique mais gagne en fiabilité (si les mesures sont bien faites), l'accoutumance diminuant les décharges émotionnelles.

Enfin, on peut encore retenir comme caractère avantageux qu'un échec ne saurait plus s'expliquer par une maladie bénigne aigüe, des circonstances météorologiques défavorables ... ou un horoscope menaçant.

Quelques inconvénients sont non moins évidents. Le premier est de surcharger considérablement le travail des enseignants. On voit,en France, des examinateurs préparer leurs questions en une demi-journée, voire une demi-heure quand ce n'est pas dix minutes. Dans la plupart des pays d'Europe (et certainement en Suède, au Danemark, en Hollande, en Grande Bretagne, en Suisse, en Belgique, en Espagne et en Allemagne), la préparation des objectifs d'examen nécessite au moins ... deux ans. Aux USA, on estime le temps nécessaire à 4 000 heures environ.. Peut être les Français ont-ils raison contre tous. Peut être peut-on admettre que tous les pays cités (et d'autres qui ne l'ont pas été, tel le Canada), ont trouvé quelques avantages à changer sans quoi on ne voit pas que, tant de temps... et d'argent aient été dépensés pour le seul plaisir d'être à la mode . Par rapport à ce qui se fait en France, la surcharge en travail à fournir est énorme; c'est vrai, mais c'est peut être utile et certainement faisable puisque celà a été fait.

Dans la mesure où des enseignants déclareraient que le travail supplémentaire est négligeable, ils seraient insconscients tout comme serait inconsciente l'autorité qui prétendrait qu'un tel système de "contrôle continu" ne nécessite aucun moyen supplémentaire (et rentable), sauf si on admet que le "CC" ne soit qu'une formalité .

Le second, le plus fréquemment souligné, c'est le temps. Non le temps nécessaire mais la situation dans le temps. En supposant un enseignement, même partiel, dans une discipline, terminé en décembre et sanctionné immédiatement par des épreuves partielles, considérant comme une fonction de la distance de l'observateur le caractère continu ou pointillé d'une ligne la question se pose de l'examen récapitulatif en juin. L'argument: vérifier que les étudiants n'ont pas oublié en juin ce qu'ils savaient (éventuellement) en décembre. Soit. Pour être logique, il faudrait encore vérifier à la fin de l'année suivante, à la fin des études et après les études, six mois représentant les deux tiers d'une année universitaire, le dixième d'un curriculum universitaire normal, un peu plus que le centième de l'espérance de vie professionnelle du candidat. Si l'enseignement avait pris fin en mai (par exemple, parce que ce n'est qu'au second semestre que des locaux se trouvaient disponibles), le Professeur aurait convenu du caractère satisfaisant de l'épreuve finale. Peut-être serait-il mieux de considérer le calendrier sous un angle pédagogique que sous celui de la cosmographie. Sauf examen de cosmographie. En vérité, on considère comme terminale une épreuve se tenant à une date fatidique (juin) et comme partielle une épreuve se tenant à une autre époque. Dans la mesure où un enseignement dure, ne serait-ce que deux ans, chacune des deux épreves subies en juin est considérée comme terminale et non comme partielle. Et pourtant ...

En allant plus loin, dans des études qui durent six ans, chaque examen de chaque discipline passé en juin est terminal (de l'année) alors qu'il n'est que partiel puisque chacun d'eux est jugé nécessaire pour parcourir convenablement le cycle des études. La vraie question n'est pas de savoir à quelle saison un étudiant doit faire ses preuves mais combien de temps il doit être capable de les fournir. Immédiatement? le partiel suffit. Dix ans après? Le contrôle tardif s'impose en plus. S'estime-t-on incapable d'exiger une épreuve en sixième année sur un sujet de première année? Quelle importance aura donc la date réelle de l'examen: novembre ou juillet? L'examen partiel et l'examen récapitulatif n'ont pas la même fonction, leur coexistence est tout à fait possible mais on ne peut pas les différencier sur leur seule situation dans le calendrier.

Le troisième inconvénient est extrêmement sérieux. La fiabilité est, toutes choses égales par ailleurs, inversement proportionnelle à la suspicion. On dit que, par fois, les épreuves jugent mieux des talents de copistes des candidats que de n'importe quel autre. Il est extrêmement dommage que tous les enseignants n'aient pas sur ce point la même attitude en dehors de toute référence politique, philosophique, sociale, religieuse ou autre. Si une épreuve doit révéler le fond propre du candidat (ce qui fait rejeter en principe l'épreuve-récitation) admettre qu'on délivre une estampille de valeur aussi bien si le contenu de l'épreuve provient du candidat, de ses voisins, de ses notes ou d'autres sources clandestines c'est admettre de faire de la fausse monnaie. C'est un mensonge. C'est remplacer une réalité par des apparences: c'est de l'hypocrisie. Pour des considérations matérielles diverses, il s'avère souvent impossible d'organiser des examens partiels dans les mêmes conditions de sécurité (très relative) qu'un examen "final". Deux solutions restent alors: rendre les communications clandestines inefficaces, ou ne pas faire subir cette épreuve qui ne prouverait rien, puisqu'on ne peut pas empêcher de telles communications qui resteraient interdites. Certes, une solide tradition, spécifiquement française dit-on, fait que les étudiants ont, dans leur immense majorité, l'habitude de passer à l'acte et ils reprochent volontiers aux enseignants de ne pas savoir les en empêcher. La vraie question est ailleurs: en nombre considérable, les enseignants n'en ont aucune envie pour des mobiles divers: ennui d'une surveillance, refus de prendre cette pratique au sérieux ou désir de ménager leur popularité. Rien de tout celà n'a le moindre rapport avec le but des études et les objectifs d'un examen.

Au total, le système des examens partiels présente des avantages qui lui sont propres si certaines conditions sont respectées et qu'on peut résumer par plus grande sûreté et plus grande exigence et la plus grande exigence est l'un des aspects essentiels de la fiabilité.

Les inconvénients ne sont pas propres au système: le gros travail, le caractère ponctuel des vérifications et l'altération des épreuves par le copiage se retrouvent aussi bien dans les épreuves "terminales" (uniques). Il est même possible d'avancer que la difficulté qu'éprouvent maints enseignants à "faire" de nombreuses questions est un avantage considérable: celà leur montre qu'ils ne savent pas ce qu'ils veulent qu'on retienne de leur enseignement puisqu'ils s'avèrent incapables de l'exprimer. Un type d'ouvrage est à cet égard révélateur: les annales d'examen. Leur lecture montre l'effarante pauvreté du choix offert aux candidats pour montrer qu'ils "savent" et ils ont bien raison de les utiliser puisque, après tout, c'est la seule source de renseignements qu'ils aient bien souvent sur ce qu'on exigera d'eux. Si les exigeances sont très limitées, ce n'est pas la faute du lecteur.

EXAMEN TERMINAL

C'est le terme habituellement utilisé pour désigner la floraison de juillet. A la fin de l'année, le jury publie la liste de ceux qui peuvent franchir un échelon. Dans des études qui durent six ans, seul l'examen terminant la sixième année mérite véritablement son qualificatif de terminal; tous les autres sont partiels puisqu'ils ne vérifient que ce qui a été précédemment enseigné et naturellement pas ce qui reste encore à étudier.

On pourrait encore concevoir un examen terminal qui serait, en plus, global. Dans l'exemple précédent, l'examen de fin de dernière année est évidemment "terminal", ce qui ne l'empêche pas d'être partiel, lui aussi, s'il ne vérifie que l'acquisition de la dernière année. S'il est estimé que tout ce qui a été enseigné et même vérifié au cours ou à la fin de chacune des années précédentes doit rester connu, compris, applicable... et qu'on ne soit pas certain que cela l'ait été, l'examen terminal doit être global.

Global est un terme habituellement utilisé à la place de récapitulatif. Un examen récapitulatif pourrait être défini comme vérifiant les choses essentielles (principales). Il y a donc plus qu'une nuance entre "terminal" et "récapitulatif". Si ce qu'on doit apprendre au cours de la première année est réputé présenter une quelconque utilité dans l'exercice prévisible de ce qui constitue le but des études, on ne voit pas pourquoi il conviendrait de fermer pudiquement les yeux sur un objet restant utile sous le prétexte qu'il a été acquis quelques années plus tôt. A moins qu'on préfère ignorer qu'il n'a pas été acquis ou que son utilité ne soit pas réelle.

Global peut encore représenter une mise en commun: une sommation. Pour savoir si un candidat est acceptable, on fait la somme de toutes ses notes et le total décide. Un tel examen de sommation est une absurdité dans la mesure où les épreuves dont on fait l'addition des notes ne sont pas de même nature. Les candidats disposent d'un système "à la carte" leur permettant de gagner "des points" dans telle discipline au détriment de telle autre. Cela signifie que le but de l'examen est de "gagner des points" et pas du tout de montrer qu'on est "suffisant" dans chaque discipline. Dans ces conditions, on admet qu'aucune discipline n'est utile. Il n'est donc pas utile d'en vérifier les acquisitions ni, peut être même, de les enseigner.

Dans la mesure où des examens partiels sont proposés et que l'unité de temps est l'année, l'examen dit terminal est l'examen de fin d'année: la question est de savoir sur quoi il va porter. A cet égard, trois solutions peuvent s'envisager:

1) Le dernier examen partiel ne correspondait pas à la dernière partie de l'enseignement (il n'était pas achevé), l'examen terminal concernera cette partie non explorée de l'enseignement: il est partiel, c'est le dernier partiel, c'est le partiel-terminal.

2) L'ensemble des partiels a exploré la totalité du champ enseigné. Le"programme" du terminal est la somme des "programmes" des partiels. On risque de se trouver dans la situation embarrassante de résultats contradictoires. Un bon terminal confirmant de bons partiels ou un mauvais terminal confirmant de mauvais partiels n'apportent aucune information utile.
Au contraire, des résultats inverses dans les partiels et dans le terminal pose la question de confiance: quels résultats doivent-ils être pris en considération? Il est extrêmement difficile de répondre. Se fier aux partiels revient à nier l'utilité du terminal qu'il n'y avait donc pas lieu de proposer. ou inversement. Prendre les dates en considération revient à dire, selon le choix, que les preuves précoces de savoir ont une plus grande importance que les preuves tardives ou inversement. Attribuer une valeur précise (un coefficient) aux résultats de chaque épreuve suppose résolue la délicate question des critères de valeur. Pourquoi attribuer 30% de la valeur globale d'une épreuve aux partiels plutôt que 35, 40 ou 29 ? En réalité une telle conception de l'examen "terminal" revient à privilégier une épreuve partielle qui a la propriété d'être la dernière. Même si les sujets terminaux sont multiples et recouvrent les sujets "partiels", l'épreuve terminale reste partielle et prend un aspect de mosaïque.

3) L'ensemble des partiels a exploré la totalité du champ enseigné. Chacun d'eux, dans un domaine restreint, indépendamment des autres. De telles épreuves sont réellement (mais relativement) ponctuelles. Chacune d'elles a permis de se rendre compte de l'état du savoir dans une partie d'une discipline. Si le découpage d'une discipline en thèmes est indispensable pour son enseignement, une telle fragmentation est une dénaturation dans la mesure où elle ne permet pas d'en saisir l'unité. Ainsi, par exemple, la "respiration" et la "circulation" représentent réellement deux disciplines différentes si on considère les objets d'étude, les méthodes et les connaissances élémentaires nécessaires à leur compréhension. Si c'est la physiologie qui est définie comme discipline (ce qui est tout aussi soutenable) une question telle que "les apports périphériques en oxygène" n'est envisageable

que sous le double aspect d'apport d'oxygène par le sang (respiration) et d'apport de sang à la périphérie (circulation). Il n'est pas évident qu'une telle question (ou beaucoup d'autres , telle la physiologie de la petite circulation) puisse être posée à aucun examen partiel parce qu'il aura été choisi de séparer l'examen de circulation de celui de respiration. Il est inéluctable, si des partiels ont été décidés de créer ce genre de domaines importants mais inexplorables, parce qu'ils portent sur deux tranches différentes. Dans ce cas il est tout à fait possible de corriger ce que peuvent avoir de ponctuels des partiels en demandant à l'épreuve terminale d'apporter la preuve d'un certain savoir qu'il n'était pas possible de demander avant. L'épreuve terminale aurait alors une fonction particulière qui justifierait aussi bien son existence que celle des partiels..

Il n'est pas inimaginable de sortir du cadre, un moment nécessaire mais étroit, des disciplines dans de telles épreuves terminales qui pourraient alors, et à elles seules, constituer un examen. Le mot "terminal" se rapporterait non à un calendrier mais bien à quelque chose d'achevé. Tout ceci suppose non seulement définis des objectifs mais encore choisi un programme d'enseignement auquel doit être asservi le programme d'examens.

DES CIRCONSTANCES PARTICULIERES

Il s'agit d'abord d'examens préalables aux études et, ensuite, des conséquences que peuvent avoir de telles opérations sur les examens liés à un enseignement.

NUMERUS CLAUSUS

La pléthore des effectifs d'étudiants en médecine inquiète, leur déflation rassurerait. Pour cela un seul moyen : la diminution du nombre à l'entrée, soit par diminution du nombre de postulants (conception optimiste) soit par voie autoritaire : c'est le numérus clausus.

Première justification du Numerus Clausus

Elle est locale. La personne investie de l'autorité sur la faculté (ou l'U. E. R.) limite le nombre d'étudiants en fonction du volume des bâtiments, de leur équipement en matériel et en personnel. Une telle personne (physique ou morale) devant être responsable, éclairée... et exister. On peut alors penser que cette limitation, agissant à la manière d'une écluse, égaliserait le niveau des biefs, étant entendu que ceux-ci n'ont pas obligatoirement tous la même surface. Ainsi, un tel N. C. serait la thérapeutique d'un trouble de la distribution.

Le résultat en serait excellent, à condition que le nombre total d'étudiants en médecine soit égal au nombre de médecins nécessaires, compte tenu d'un rendement normal des études.

Deuxième justification du Numerus Clausus

Elle est nationale et peut être plus large compte tenu des termes du traité de Rome sur la libre circulation des "travailleurs" pour peu que les médecins soient considérés comme tels.

A supposer que le nombre d'étudiants pouvant être raisonnablement admis dans les U. E. R. soit inférieur aux besoins, le choix qu'il faudrait faire entre le sous équipement médical et l'équipement universitaire convenable serait politique et non plus technique. Si, par impossible, le sous équipement universitaire était choisi, le traité de Rome aidant, le sous équipement médical ne serait pas inéluctable d'autant qu'on trouve probablement plus de Hollandaits parlant le Français que de Français parlant flamand.

A l'inverse, il serait très dommage que le nombre de médecins formés soit supérieur à celui que peut absorber le "marché" : tant pour les intéressés eux-mêmes que pour la collectivité qui finance les établissements d'enseignement.

Qu'un tel N. C. ne verse ni dans l'excés ni dans le défaut implique une connaissance des besoins (au moins à moyen terme, les études durant six ans au minimum) et de la couverture probable de ces besoins dans ces délais.

Connaître "la couverture" ce n'est pas seulement connaître le nombre des docteurs en médecine. Il faut savoir combien sont actifs, prévoir la durée d'activité et même préciser ce terme. En dehors des médecins à la retraite, il faudrait pouvoir éliminer des comptes ceux qui sont inscrits au tableau de l'ordre sans exercer à proprement parler : enseignants, fonctionnaires, chercheurs, employés dans des laboratoires... etc, sans omettre les docteurs qui exercent l'honorable profession de mère de famille. Or il semble que, dans cette numération, les données sont très approximatives.

Connaître les besoins semble encore plus difficile pour le présent même, à fortiori en ce qui concerne l'avenir, la répartition des spécialistes en spécialités actuelles et futures et celle des spécialités actuelles dans le futur.

Au total, un N. C. peut être justifié à l'échelon local, régional, national ou international, en tant que moyen d'adapter un volume à une capacité. Mais un N. C. étant un nombre, encore faut-il que les données soient chiffrées et que les grandeurs qu'il s'agit de comparer deviennent mesurables et soient mesurées.

X X X
X X
X

LA SELECTION

C'est un choix raisonnable. En soi, elle n'a aucun rapport de nature avec le N. C. Il s'agit, en fait, de juger un postulan "apte" ou "inapte". Cela suppose donc que l'aptitude à être docteur en médecine soit définie.

Ce peut être l'aptitude aux études supérieures : c'est la définition même du baccalauréat. On peut envisager que les études de médecine sont non pas des études supérieures, mais l'apprentissage d'une profession (enseignement professionnel). Dans ce cas, il resterait à déterminer les qualités requises pour l'exercice de la profession : physiques, mentales, intellectuelles et peut être même, morales. En supposant connu le "profil" du médecin, une sélection peut être opérée. L'est-il ? Et le serait-il, qu'il faudrait encore prévoir une évolution individuellle qui ne semble pas montrueuse, notamment chez des sujets jeunes.

Il est aussi difficile de nier l'aspect "supérieur" que l'aspect "professionnel" des études de médecine. Séparer ces deux aspects n'est

peut être pas souhaitable mais n'est possible que si, dès l'entrée, deux filières sont prévues : celle des "praticiens" et celle des "savants" qu'on peut encore appeler celle des contre maîtres et celle des ingénieurs. Les cycles ne seraient plus successifs mais parallèles et des "passerelles" toujours imaginables seraient peu pratiquées. Un tronc commun initial pourrait, certes, retarder le moment du choix mais pour être rentable un tel tronc devrait être court et pour être sûr il devrait être long. Enfin, que "savants" et "praticiens" soient longtemps confondus ou d'emblée séparés, il semblerait logique qu'une sélection s'opère encore au début de la spécialisation ; moins dramatique pour un candidat dont le changement d'orientation est relativement aisé à ce stade, il ne faudrait pas qu'elle constate "à posteriori" une simple aptitude aux études médicales.

SELECTION ET NUMERUS CLAUSUS

Cette combinaison donne un troisième phénomène tout comme un circuit électrique comportant une résistance et une capacité est différent de celui qui ne comporterait que l'un ou l'autre. Il ne s'agit plus de sélectionner les sujets aptes mais les meilleurs. Cela nécessite l'établissement d'échelles de valeurs précises devant être le plus discriminatives là où elles ont le moins de chances de l'être : vers la moyenne. S'il en était autrement, on accorderait une signification au centième de point là où deux points n'ont pas de signification. D'autre part, le numerus clausus doit être rigoureux : rigoureusement établi, tenant compte de toute la ventilation des effectifs au cours des années d'études. Qu'adviendrait-il, par exemple, d'une fraction des effectifs admis pour être anesthésiste, si dix ans plus tard, la profession est saturée ? On pourrait alors s'apercevoir qu'on manque de psychiatres. Ainsi des sujets sélectionnés à priori comme anesthésistes deviendraient psychiatres. De bons psychiatres ? Alors la sélection serait mauvaise. De mauvais psychiatres ? La sélection aurait été juste mais serait-ce une consolation pour les malades ?

EXAMEN D'UN EFFECTIF SELECTIONNE

La seule question qui se pose est de savoir si on peut ou non confondre aptitude et capacité. La seule réponse admissible est "non".

Dans la meilleure des hypothèses, un médecin serait tellement bien défini en termes d'aptitudes qu'aucun étudiant ne serait sélectionné s'il n'était pas apte. Le numerus clausus aidant, et toujours dans la meilleure des hypothèses, un nombre non négligeable de sujets aptes auraient été éliminés ; seuls les plus aptes resteraient. Pour se faire une idée du caractère utopique actuel d'une telle hypothèse, il suffit de demander à dix enseignants d'un rang de responsabilité élevé (et de compétence) quelle est la principale qualité d'un médecin : celle dont l'absence constatée serait rédhibitoire. Il n'y aurait pas autant de réponses que de sujets interrogés mais la réponse faite

le plus souvent sera "je ne sais pas". Ce test peut être fait par n'importe qui.

Quoi qu'il en soit, une faculté reçoit un certain contingent d'étudiants "aptes" ou "très aptes" dont il s'agit de faire des médecins en six ans. Ils sont aptes aux études de médecine : ils possèdent non seulement des aptitudes intellectuelles mais encore des aptitudes "humaines". Ils ont tout ce qu'il faut pour devenir médecins. Ils ne sont pas capables d'être médecins : s'ils en avaient toutes les qualités de façon innée, les études, à supposer qu'elles soient maintenues, ne seraient qu'une intolérable brimade. Le choix qui s'imposerait le jour de l'inscription en faculté serait le suivant : attribuera-t-on sur le champ un diplôme prenant effet six ans plus tard ? Si on pense qu'une telle attribution est prématurée, cela signifie-t-il qu'il est choquant de signer un papier officiel dont la prise de valeur est aussi tardive ou bien qu'il n'est pas certain que finalement tous ces étudiants aptes seront des médecins capables ?

Si on conserve un doute sur le résultat final, exprimé en termes de capacités, il faut que ce résultat soit établi sur une (ou des) épreuve de capacité. Aussi parfaitement sélectionné qu'ait été un effectif, l'examen ne perd rien de son utilité. On peut même soutenir que son exigence n'en est que plus grande puisqu'il ne lui restera, enfin, qu'à prouver ce qu'il doit prouver : une capacité.

Dans une perspective d'avenir à très court terme, il n'existera plus aucun alibi pour les examens aux résultats incertains. Il ne sera plus possible d'invoquer le niveau lamentable des étudiants pour justifier des résultats surprenants : le niveau aura été choisi. Les étudiants auront, en cas d'échec, l'impression que le contrat aura été violé : à leurs yeux, entrer dans une faculté de médecine sera une garantie d'en sortir médecin. Des enseignants "recevant tout le monde" risqueraient de recevoir n'importe qui. Pour refuser, il faudra connaître le motif. Tout échec pourra être considéré comme le fait de l'enseignement. Beaucoup de "flou" commode entourant un examen ayant disparu, il ne suffira plus de recevoir ou d'accepter, il faudra encore savoir pourquoi à un très haut niveau de précision.

ENSEIGNEMENT ET EXAMEN

Si apprendre c'est acquérir quelque chose de plus, dans le domaine cognitif ou dans un autre, l'examen a pour fonction de certifier que le but est atteint. L'enseignement c'est l'ensemble des moyens employés pour parvenir au but. La fonction de l'enseignement n'est pas de décrypter le catalogue de ce qu'il faut acquérir. On peut concevoir des rapports tout à fait différents entre enseignement et examen selon l'idée qu'on se fait de l'enseignement.

Qu'arriverait-il si, pure hypothèse, une population d'enseignés refusait toute activité scolaire ou universitaire ? D'abord, il faudrait faire la part de l'exagération des mots : entre ne pas fréquenter (acte ostensible) les activités réglementairement prévues et n'exercer aucune activité, il y a plus qu'une nuance. Bien entendu, le grade ne pourrait être délivré qu'après contrôle. Sur ce point, trois attitudes peuvent être imaginées.

1° Considérant le fait qu'aucune activité officielle ne s'est manifestée, on peut en tenir compte et "abaisser" le niveau de la vérification au niveau supposé de la population. Logiquement, il faudrait égaliser la valeur de ce contrôle et celle de la vérification précédente. Un tel examen n'aurait évidemment aucune valeur.

2° Considérant que, si aucune activité ne s'est tenue, c'est par mauvaise volonté des candidats, on peut en "tenir compte" et rédiger une épreuve telle qu'elle soit inattaquable sur le plan de la forme et insoluble sur celui du fond. Cela supposerait que le seul point critiquable d'un examen soit sa forme. Un tel examen n'aurait pas davantage de valeur que le précédent.

3° Beaucoup plus intéressante est l'hypothèse dans laquelle les objectifs d'un examen seraient définis d'une telle manière que personne n'aurait besoin de guide pour les interpréter. C'est alors sur leur valeur réelle que seraient jugés les candidats. Si tout le monde est reçu, c'est que l'enseignement ne sert à rien. Si personne n'est reçu c'est que l'enseignement est indispensable. Sans aucun enseignement, certains seraient reçus : ceux-là auraient fait la preuve qu'ils n'en avaient pas besoin.

De tels candidats montreraient ce qu'est réellement l'enseignement Le but est atteint, c'est donc que les moyens propres ont été employés par les étudiants, eux-mêmes qui sont leur véritable enseignant. Un nombre, peut être considérable, d'étudiants ne sauraient pas parvenir au but sans aide. L'enseignant de métier n'est qu'un aide. Il aime à croire que son aide est précieuse et il semble qu'il ait raison mais son importance n'est certainement pas la même vis à vis de tous les individus constituant l'effectif dont il est chargé. Ne serait-ce pas, même un peu, pour se valoriser que tant d'enseignants répugnent à ce que les objectifs soient définis ?

Si eux seuls ont à la fois le droit d'interprétation et la maîtrise totale des examens, ils sont évidemment indispensables. Dans la mesure où chacun interprète à sa manière, en fonction de ses goûts, de ses compétences ou de ses intérêts - ce qu'il convient de savoir finalement - on risque de fixer des objectifs aussi différents les uns des autres que le sont les enseignants entre eux à des étudiants qui sont réputés parvenir au même but. Dans l'état actuel des choses, il n'y a pas de risques de disparité: la disparité est un fait.

Si, au contraire de ce qui se fait, le but final, éventuellement quelques buts intermédiaires, sont clairement définis, la tâche consiste à les atteindre. Ce sont les étudiants qui doivent les atteindre et les enseignants qui doivent les y aider. L'aide consiste notamment à établir un programme (un plan d'action) consistant d'abord à organiser les objectifs et ensuite à choisir les moyens les mieux appropriés (méthodes pour les atteindre). Le choix fait d'un programme et de méthodes ne peut pas être, a priori, considéré comme le meilleur possible . On peut parier qu'aussi bon soit le choix, il est encore perfectible. La qualité peut être encore meilleure. Le temps mis à atteindre un certain niveau, réduit. En Europe, l'une des caractéristiques les plus communes des routes, c'est que les panneaux indicateurs sont situés après les carrefours et, de préférence, masqués par un arbre, un poteau, un autre panneau ou un quelconque obstacle. Le carrefour franchi, le panneau peut aussi bien annoncer "vous êtes sur la bonne voie" que "vous vous êtes trompés". C'est bien ainsi que se présentent beaucoup d'examens, même partiels. Il ne semble pourtant pas impossible d'utiliser des renseignements fournis par l'examen pour modifier l'enseignement au cas où le trajet suivi ne s'avèrerait pas aussi bon que prévu.

Tout ce qui précède peut ainsi se résumer: un examen (terminal) atteste que le but est ou non atteint à la suite d'un enseignement ou sans enseignement préalable; une épreuve (partielle) atteste qu'une étape a été atteinte, ce qui suppose qu'un itinéraire avait été fixé et c'est déjà un fait d'enseignement. Elle peut encore évaluer l'adaptation des moyens choisis aux buts définis: elle contrôle l'enseignement. S'il semble souhaitable qu'un enseignement prépare à un examen, tout enseignement doit-il préparer à un examen ?

Supposons que dans un établissement d'enseignement supérieur, une faculté de médecine par exemple, il soit considéré comme souhaitable que les étudiants sachent lire et même parler et comprendre l'anglais ou n'importe quelle autre langue. Il est évident qu'on peut être bon Médecin sans connaître un seul mot d'anglais. Il est clair que les publications

jugées intéressantes dans une perspective d'enseignement peuvent être traduites par un service de traduction. On ne peut pas déclarer indispensable la connaissance de l'anglais pour l'ensemble de la population à ce niveau (au niveau des études secondaires et surtout primaires, les critères seraient différents et le jugement pourrait n'être pas le même). Un enseignement peut être organisé. Il est évidemment absurde, si un examen est décidé en fin de première année par exemple, de compenser une insuffisante connaissance de la physiologie par une note excellente en anglais. Si la connaissance d'une langue est jugée "utile" et non "indispensable", son étude ne doit pas être sanctionnée par un examen puisque, dans les études considérées, il serait absurde de "bloquer" un étudiant insuffisant tout autant que de favoriser le sujet exceptionnellement brillant dans cette seule discipline. Il serait d'ailleurs illogique d'imposer une telle étude mais il semblerait excellent de la proposer. On pourrait choisir de l'entreprendre ou non: ce serait un enseignement"optionnel" dont on peut souligner au passage que de nombreux autres types pourraient être définis sur un critère d'utilité.

A aucun moment du curriculum le jury n'aurait à connaître de la valeur des candidats dans ce type d'enseignement, ni même du fait que tel étudiant a choisi d'en bénéficier ou non sans quoi on ne pourrait plus parler de choix. Cependant, l'établissement d'enseignement aura chargé une personne (ou une équipe) d'assurer cet enseignement et l'aura dotée des moyens appropriés. Les étudiants, tout autant que les enseignants, auront besoin de faire le point: quels sont les besoins? (pourquoi apprendre la langue? pour lire? écrire? comprendre la parole? parler? faire un stage à l'étranger?...) quel est le niveau de départ? quels sont les progrès faits? Le but choisi (lire ou parler) est-il atteint? les moyens utilisés sont-ils réellement appropriés? Il est de l'intérêt de tout le monde qu'il puisse être répondu à ces questions. Il semble bien que ce qu'on ait inventé de mieux pour faire ce genre de vérifications, ce soit une série d'épreuves. Le résultat n'a pas de valeur légale, ce qui signifie qu'il n'est pas établi une définition du niveau des exigeances permettant l'attribution d'une estampille officielle. Cependant, l'indication du niveau atteint (ce qui suppose qu'il soit défini!) constitue une information utile tant pour l'étudiant que pour les enseignants. On ne peut donc pas déclarer inutile l'examen, même pour un enseignement facultatif, sauf si on se contente d'agir sans but et sans souci de l'efficacité de l'action, ou qu'on se satisfasse de n'importe quel signe pour prouver n'importe quoi!

L'enseignement obligatoire, comme l'enseignement facultatif bénéficient des mêmes sortes d'épreuves qui ont, finalemnt, la même fonction

prouver que le but choisi est atteint ou non et prévoir que l'étudiant saura exécuter ce qu'il prétend. La seule différence, et elle est considérable c'est qu'on ne peut pas s'opposer à ce qu'un individu sachant dix mots et une phrase s'embarque pour les Etats-Unis, mais on doit s'opposer à ce qu'un individu incapable exerce la Médecine .

Au total, l'enseignement même optionnel, est fait dans un but qui doit être défini et atteint. La seule manière de savoir où on en est, c'est d'organiser une épreuve: l'examen est non seulement toujours utile mais toujours indispensable. Il est indispensable, en vue de l'exercice de certaines fonctions, d'acquérir un certain savoir alors que telle autre acquisition est simplement utile. Ce n'est pas l'examen qui change, mais la valeur qu'on y attache par ce qu'il garantit. Certains examens, par le contenu de ce qu'ils attestent - leur cahier des charges en quelque sorte - sont des diplômes. Il ne semble pas qu'il y ait motif d'instituer, pour des médecins, des diplômes d'anglais, de dactylographie ou de sténographie , bien qu'on ne puisse nier l'utilité de telles acquisitions.

UN PLAN D'EXAMEN

Soit à concevoir un examen de "clinique" à un niveau élevé des études de médecine. Il sera arbitrairement choisi de considérer que le contenu d'un tel examen devra concerner la sémiologie, la pathologie et la thérapeutique.

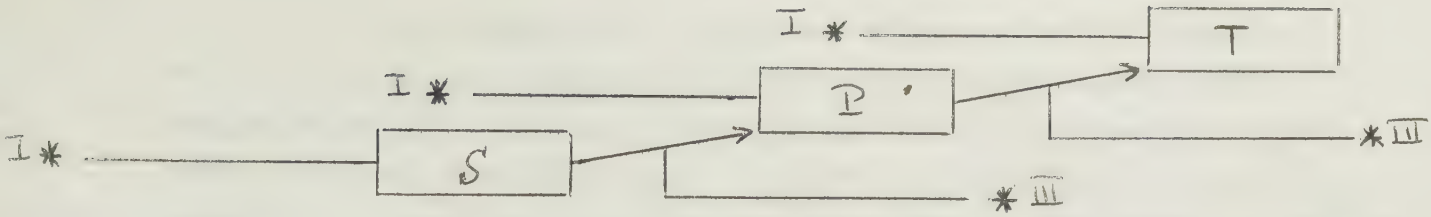
Il est très difficile d'établir un plan si on ne sait pas comment a été conçu l'enseignement, de même que l'enseignement fait sans préoccupation de ce que sera le contrôle risque d'être insuffisant ou mal centré, à condition, bien entendu, que chacune de ces activités soient orientées dans la même direction. Des trois éléments considérés, on peut raisonnablement faire deux ensembles :

Une épreuve de pathologie (P) au sujet de laquelle seront demandées des preuves de savoir en sémiologie (S) et en thérapeutique (T).

Une épreuve de thérapeutique dans laquelle seront demandées des preuves de savoir concernant le tableau de la maladie (P), la signification des signes (S) et la manière de traîter.

Quel que soit le schéma retenu, l'objectif est insuffisamment décrit pour peu qu'on souhaite dépasser le simple "contrôle" des connaissances" . Si le premier schéma est retenu, on peut ainsi le résumer "décrire le tableau de telle maladie, en précisant la technique de recherche et la signification des signes qui le constituent et proposer un schéma de traitement". Cela peut revenir à la constitution d'une fiche pour chaque maladie, éventuellement très résumée, qu'il n'y a plus qu'à réciter.

Si le mode d'enseignement choisi a été le "bloc teaching" par exemple par bloc court, le schéma, très cartésien, en a été le suivant:



A partir d'éléments simples (sémiologie), on construit un ensemble plus complexe (le tableau pathologique), puis on indique les moyens d'action thérapeutique. Plusieurs contrôles sont possibles (✱) les uns de connaissances (I) , les autres d'application (III). En effet, il est possible de demander la description de la technique de recherche des signes rencontrés dans la maladie, ce qui suppose qu'on les connaisse (I) et qu'on sache choisir (III), ce qui est utile ou utilisable parmi ce qui est possible et de même pour le choix des éléments de traitement. Ainsi cinq épreuves partielles peuvent être conçues dont la somme peut constituer "un certificat" dont on admettra ou non que ses résultats puissent être remis en question.

Une telle épreuve est-elle probante? Cela dépend de ce qu'on veut prouver. S'agit-il de connaître les signes, de savoir les décrire eux-mêmes et leur technique de recherche et d'indiquer un schéma de traitement? Elle est probante ou peut l'être si elle est bien constituée. S'agit-il de montrer qu'en présence de cet ensemble de signes on saurait reconnaître la maladie? Il est simplement plus probable que soit identifié quelque chose de connu. L'épreuve n'est pas probante. L'opération consistant à prescrire à partir d'un malade donné est beaucoup plus complexe (même si le malade est "simple") et beaucoup des actes élémentaires indispensables n'ont pas été explorés: on ne peut donc être sûr que le candidat soit capable de les accomplir .

Tout commence, typiquement, par un malade qui se plaint de quelque chose. Il n'est pas habituel que le patient s'exprime en un langage techniquement correct: il faut le comprendre (II), traduire ses déclarations en termes de signes fonctionnels qu'il faut connaître (I) et déjà grouper dans un tableau sommaire mais souvent évocateur (IV). Entre toutes les techniques possibles de recherche de tous les signes physiques possibles qu'il faut connaître (I), il s'agit de choisir (III) (application), puis comprendre (II) ce qu'on a trouvé pour l'ajouter au tableau précédent qui devient moins sommaire (IV). Il convient de procéder de même dans la quête des informations complémentaires en négligeant les innombrables raisons qu'on peut avoir de passer de l'examen physique ou des examens complémentaires à l'interrogatoire, à partir de l'ensemble peu à peu constitué.

Cet ensemble est le résultat d'une analyse (IV) complète puisqu'il s'agit non seulement de rechercher les éléments mais encore quelles relations existent entre eux et quel principe d'organisation les relie. Tout ceci doit être comparé à toutes sortes de connaissances pour aboutir (V : constitution de relations abstraites) à deux choses: le diagnostic et un doute. Le diagnostic établi doit être comparé à la connaissance de différents tableaux pathologiques (diagnostic différentiel) dans un but de critique (VI), ce qui a pour conséquence de diminuer plus ou moins le doute .

A partir du diagnostic plus ou moins douteux, on établit un "plan d'action" qui est un schéma thérapeutique (V)en omettant les circonstances où la critique impose de recommencer tout ce qui précède. Le traitement est susceptible de modifier le tableau pathologique (ce qu'il faut savoir rechercher, comprendre, analyser et comparer à ses connaissances) et, dans les bons cas, les modifications observées diminuent encore le doute: elles constituent la meilleure des critiques externes. Ainsi peut être schématisée la démarche intellectuelle aboutissant à l'acte médical, à côté de laquelle un simple contrôle des connaissances fait bien misérable figure.

Faut-il tout contrôler? La réponse est affaire de conscience. En fait, autoriser quelqu'un à exercer la médecine, c'est certifier qu'il est capable de faire tout cela dans de nombreux domaines. Il appartient à ceux qui ont l'honneur et la responsabilité d'une telle charge de se prononcer sur la valeur de leurs propres affirmations.

CHAPITRE XI

FORMES D'UN EXAMEN

E - CHOISIR UNE FORME ADEQUATE

" ... entendez les parler, les plus habiles gens du monde; voyez les faire, les plus ignorants de tous les hommes ".

MOLIERE.

QUELQUES IDEES PRELIMINAIRES 353
UN TRES MAUVAIS EXAMEN DE MEDECINE 355
FORMES TRADITIONNELLES 360
DES FORMES MODERNES 363
NECESSITE D'EXAMENS NON VERBAUX..... 370
PERSPECTIVES D'AVENIR 371
ANNEXES 379

QUELQUES IDEES PRELIMINAIRES

1) La finalité d'une épreuve est la manifestation d'une preuve. La nature de la preuve dont on souhaite la manifestation doit être connue. C'est en fonction de la preuve recherchée que doit être construite l'épreuve.

2) Plus généralement, le rôle d'une épreuve est d'attester qu'un objectif a été atteint ou d'évaluer dans quelle mesure il a été atteint.

3) L'ensemble des objectifs à atteindre dans un curriculum constitue ce qu'on espère représenter la valeur ajoutée, finalité de l'enseignement. L'organisation de ces objectifs en fonction du temps mérite le nom de "programme".

4) La finalité des études, celle de l'enseignement et celle de l'épreuve doivent être les mêmes. La différence essentielle entre les objectifs des études et ceux de l'enseignement, c'est que les premiers forment un tout défini en termes de masse organisé, les seconds sont définis en termes de progression. L'épreuve ne doit naturellement pas être conçue de la même manière si elle vérifie une progression ou l'atteinte du but final.

5) C'est par rapport au niveau "actuel" souhaité et non au niveau "terminal" espéré que doit être jugée la valeur démonstrative d'une épreuve non terminale.

6) Dans la détermination du niveau souhaité des étudiants, une méfiance systématique doit s'exercer vis-à-vis du juge. Dans un curriculum normal d'enseignement supérieur, 1er et 2ème cycle, l'objectif est de former des "praticiens généralistes". Or, les professeurs sont habituellement des "chercheurs spécialistes" et leurs propres canons de valeurs ne s'appliquent pas, en général, à la nature du produit fini souhaité.

Soient, par exemple, les études de médecine, considérées au cours du premier cycle. Des disciplines fondamentales y sont enseignées : c'est un fait. L'utilité de telles sciences n'est pas contestable et assez peu contestée. Deux questions se posent :

a) est-il souhaitable que l'ensemble de telles connaissances soient acquises au cours du premier cycle ?

b) les critères de valeur de l'équipe enseignante doivent-ils être systématiquement acceptés ?

Le chercheur spécialiste possède une connaissance caractérisée par sa profondeur et sa haute technicité. Il est au fait, dans sa spécialité, de l'état actuel de la question, mais une quantité considérable d'informations qu'il possède sont actuellement inexploitables pour le praticien d'une autre discipline dont la culture dans le domaine est insuffisante, pour qui l'application est actuellement impossible et que les techniques particulières à une discipline qui n'est pas réellement la sienne n'intéressent pas.

Au total, bien souvent, on pourrait accuser les enseignants de manquer de réalisme en supposant universellement intéressant ce qui n'intéresse qu'eux, de manquer d'objectivité en ignorant systématiquement les caractéristiques des étudiants dont ils ont provisoirement la charge et de craindre le ridicule en n'exigeant pas des connaissances qui donnent l'apparence d'un haut niveau d'enseignant. Ce qui est ridicule, c'est pourtant la possession d'une forme sans fond.

UN TRES MAUVAIS EXAMEN DE MEDECINE

Il s'agit du plus prestigieux : la thèse. Sa définition peut ainsi se résumer "un mémoire d'une longueur quelconque sur un sujet quelconque". L'acceptation entraîne l'attribution du grade de docteur, et c'est logique, et l'autorisation d'exercer la médecine, ce qui ne l'est pas. Fondamentalement, la thèse est un discours. Quelle qu'en soit la forme, brillante ou non, sérieuse ou superficielle, lourde ou légère, brève ou longue, elle ne révèle à ce titre, que le talent de l'auteur à exprimer ses idées, à supposer que les idées exposées soient les siennes, ce qui n'a rien d'obligatoire. Quant au fond, il est tout à fait admissible de devenir docteur en médecine en exposant un sujet de docimologie, ce qui ne garantit en rien la moindre compétence dans le domaine particulier du médecin,qui est d'établir un diagnostic et de prescrire une thérapeutique. On peut évidemment supposer que, sur ce sujet, d'autres contrôles ont été faits. Mais alors à quoi sert la thèse ?

1) Ce n'est qu'une formalité, puisque ni la forme, ni, surtout, le fond n'en sont définis. La sanction en est un certificat attestant que M. Untel a soutenu sa thèse à telle date, ce qui permet de s'inscrire à l'Ordre et d'avoir une existence professionnelle légale. Ainsi considérée, elle ne prouve rien et ne sert à rien d'autre que d'accomplir une obligation d'utilité douteuse.

2) C'est un exercice de haut niveau intellectuel : une synthèse (V) qui peut se décrire comme une communication originale dans le but de décrire, d'informer, de persuader et peut-être même d'émouvoir(mettre en mouvement). Synthèse dont la réalisation exige des connaissances (I) et de la compréhension (II), dont la conception nécessite une analyse (IV) si on part de "faits", de l'application (III) si on propose des solutions et de la critique (VI) si on a un minimum d'amour propre. C'est un exposé censé apporter des informations jusqu'alors peu ou mal connues. C'est un travail d'enseignement dont le succès est sanctionné par le titre de Docteur "celui qui enseigne" (docere) . A ce titre, une thèse prouve quelque chose dans la mesure où on la juge sur les critères de jugement d'une synthèse : la liberté de pensée et d'action ont-elles été totales, sans point de vue ni conclusion imposée ? Le temps suffisant pour sa rédaction a peut-être été long, ce qui impose la nécessité de se contenter d'une seule production qui sera finalement jugée plus sur la manière de parvenir aux conclusions que sur les conclusions elles-mêmes.

Un aspirant docteur en médecine qui aurait travaillé sur un thème d'enseignement n'aurait pas nécessairement à craindre d'avoir perdu son temps s'il considère qu'il ne s'est pas pour autant éloigné de la profession qu'il a choisie en l'envisageant sous l'un de ses innombrables aspects originaux. Il ne lui reste que sa conscience pour savoir qu'un tel travail ne lui a guère permis d'approfondir ses connaissances en thérapeutique. L'exercice consistant à faire une communication(un discours) n'a rien de choquant dans la mesure où le genre d'activité souhaitable est de faire des communications. Il est très mauvais, en revanche, de se contenter de faire un discours sur les maladies si l'exercice prévisible consiste à les reconnaître et à ce titre, l'"esprit thèse" est très mauvais en médecine.

Chacun sait qu'un chirurgien est un homme habile de ses mains. Certes, un bon chirurgien n'est pas seulement un bon opérateur, mais on peut raisonnablement espérer qu'il est, notamment, un bon opérateur. Or, il n'est pas demandé officiellement à un aspirant chirurgien d'apporter la preuve qu'il sait opérer ! Voici comment il doit faire ses preuves(le texte authentique en annexe).

. Deux exposés écrits dont l'un porte sur la pathologie et l'autre sur la technique. Chacun de ces exposés est noté sur vingt et il est exigé d'obtenir au moins 20 à l'ensemble pour subir l'épreuve pratique.

. Qui consiste en l'examen d'un malade(un examen de malade au singulier) pendant une heure sans qu'il soit le moins du monde indiqué à quoi doit aboutir cet exercice dont on sait qu'il dure une heure !

Ainsi la seule preuve exigée d'un chirurgien pour attester qu'il sait opérer est un discours sur la technique chirurgicale dont on suppose qu'il saura l'exécuter s'il sait la décrire. Encore est-il possible qu'un parfait discours sur la pathologie le dispense de savoir quelque chose sur la technique. On peut évidemment supposer que l'apprenti apprend réellement quelque chose sur son lieu d'apprentissage et on peut espérer que le jugement du "patron" aura finalement quelque importance. A quoi donc sert cet examen ?

Il n'est pas demandé à accoucheur de montrer qu'il sait faire un accouchement. Après avoir réussi deux à quatre discours préliminaires, le prétendant obstétricien est admis à subir l'épreuve pratique suivante :

... /...

"... examen en quinze minutes d'une femme enceinte ou d'une parturiente, ou d'une accouchée, ou d'un nouveau-né, suivi, après dix minutes de réflexion, d'un exposé oral(encore un discours) de dix minutes, puis d'une interrogation portant sur une ou plusieurs des matières suivantes...". Epreuve notée de 0 à 20 d'importance égale à l'épreuve de thérapeutique qui n'est absolument pas définie. Ainsi, les seules précisions sur l'examen de fin d'études du CES d'obstétrique sont que :

1) on doit examiner une femme ou un enfant pendant quinze minutes.

2) réfléchir dix minutes.

3) discourir dix minutes.

Après ces trente cinq minutes qui comptent dans la vie d'un homme, on subit une interrogation sur un ensemble hasardeux dans des formes imprécises. Pour des raisons qu'on espère bonnes, le jury, composé d'au moins cinq membres, est désigné par le Ministre de l'Education Nationale. Ces juges n'ont à juger que de la première partie de l'examen terminal. L'épreuve "pratique" est jugée par des membres de la faculté à laquelle appartient le candidat, ce qui peut laisser supposer que s'instaurent certaines traditions à l'intérieur de chaque faculté, mais ne préjuge en rien de la moindre similitude entre les traditions respectives de chaque faculté qui, chacune de son côté, "estime" ce qu'il doit être nécessaire de savoir pour être accoucheur. Bien entendu, exiger un accouchement est illégal.

On peut encore réfléchir sur ce qu'est un radiologue avec d'autant plus de sérénité que le certificat considéré a cessé d'être délivré à partir du 31 décembre 1971. Il faut d'abord écrire trois discours en trois heures(après une demi-heure de réflexion) sur les sujets suivants :

a) radiodiagnostic (maximum 30)

b) roentgenthérapie ou curiethérapie (maximum 30)

c) électro-physiothérapie (maximum 20)

Total maximum : 80

... /...

Un candidat ayant obtenu 40 (la combinaison 0 + 20 + 20 n'est pas exclue) subit une seconde partie.

a) discussion d'un mémoire (maximum 20) dont on sait qu'il doit être rédigé en 5 exemplaires, mais dont on ignore le genre de sujet.

b) une épreuve clinique : examen d'un malade, prise ou lecture de clichés (maximum 20). Il serait peut-être de mauvaise foi de prétendre que ces clichés peuvent être photographiques, mais il est incontestable que la radioscopie est exclue.

c) une épreuve orale (encore un discours) portant sur l'ensemble du programme (maximum 20). A noter qu'il serait illégal de faire porter l'épreuve sur un sujet de neuroradiologie qui ne figure pas au "programme" !

Au total, 140 points peuvent être attribués dont 20 au maximum à l'examen d'un malade, à la prise ou à la lecture de clichés, soit un septième ! La partie cachée de l'iceberg peut être ainsi constituée :

1ère partie :

a) Protection contre les dangers des rayons X utilisés en radiodiagnostic : principales dispositions légales actuellement en vigueur(30)

b) Moyens de protection contre les dangers des rayons X utilisés en roentgenthérapie : principales prescriptions légales actuellement en vigueur (20)

c) Législation actuellement en vigueur au sujet des générateurs de rayons infrarouges utilisés en thérapeutique(20).

2ème partie :

a) appréciation d'un mémoire portant sur "la lumière considérée comme agent thérapeutique de l'Antiquité à nos jours. Perspectives d'avenir, incidence sur le taux de natalité"(20).

b) lecture des clichés, en vue du diagnostic, d'un sujet atteint de fracture de Dupuytren dans sa forme typique non compliquée, sans l'évolution(20).

c) Enumérez cinq corps radioactifs en précisant pour chacun d'eux sa masse atomique(20).

... /...

Incontestablement, un candidat ayant réussi à répondre brillamment à chacune de ces questions saurait quelque chose. On peut cependant avancer l'hypothèse qu'il n'est pas nécessaire de faire trois ans d'études supplémentaires pour répondre à la deuxième partie dans sa subdivision "b" et que, plus généralement, ce qui est su n'a pas la moindre importance si on considère d'une part que tout cela a le maximum de chances d'être vite oublié et d'autre part qu'on attend généralement d'un radiologue qu'il sache faire et comprendre une radio dans n'importe quelle situation pour aboutir à n'importe quel diagnostic(pourvu qu'il soit juste). Si un radiologue ne peut pas, ne sait pas faire face à ce genre de circonstances, personne n'en a besoin.

Au total, se contenter de discours à titre de preuve qu'on sait agir ne peut en aucun cas être considéré comme satisfaisant. Un médecin doit agir avec des moyens concrets sur des objets concrets. Qu'il soit capable de faire des discours savants prouve peut-être qu'il est savant, mais non pas qu'il est efficace et c'est pourtant ce qu'on attend de lui tout aussi bien que d'un enseignant. C'est là le véritable danger de ce qui s'est appelé plus haut l'"esprit thèse" et qui situe à un niveau relativement modeste les querelles entre les partisans des "examens rédigés" et ceux des "questions à choix multiples" ou des "examens programmés".

FORMES TRADITIONNELLES

Que le discours demandé soit oral ou écrit ne change rien au fond de la question : que prouve la réponse ? Tout au plus est-il possible, à ce sujet, de souligner quelques contradictions fondamentales semblant appartenir au patrimoine commun de tous les examinateurs. Sous une forme ou sous une autre, il est possible de demander un long discours(tels ceux qui ont été envisagés dans les certificats d'études spéciales) ou un petit ou plusieurs petits discours. Cela suppose en fait que la mélodie du choeur des examinateurs commence toujours par une variation autour du thème "que savez-vous de ...".

Il peut arriver que la question soit sans réponse(cela s'est vu au concours de l'Internat des Hôpitaux de Paris) ou que la réponse soit clairement soufflée, ou que la réponse soit nécessairement sans signification (pourquoi Jupiter se trouve-t-elle entre Mars et Saturne ?) ou que la réponse exigée soit arbitraire(quels sont les deux principaux nerfs du membre supérieur, etc...)

Autant d'écueils qui pourraient être évités si, à partir de la réponse attendue, on rédigeait la question : les arrières pensées des rédacteurs ne sont pas toujours évidentes et les découvrir dépendent plus d'un mécanisme de jeu du hasard que du processus rationnel qu'on demande pourtant aux candidats.

On accorde volontiers à cette forme l'avantage exclusif de permettre de juger des moyens d'expression du candidat. Il est vrai, que seule une forme d'épreuve exigeant une expression verbale permet de juger de cette expression. Trois conditions sont requises pour qu'un tel jugement puisse s'exercer.

1) Que l'examinateur ne confonde pas "agréable à lire" et "bien exprimé". Chacun, sans l'avouer peut-être, préfère la lecture de tel auteur de roman policier à celle de tel auteur monumental, à l'expression parfaite, mais ennuyeuse.

2) Que l'examinateur soit certain que le texte lu est bien de l'auteur supposé et non, par exemple, du professeur(qui pourrait lui-même être ennuyeux) que ce soient les moyens qu'on juge et non l'expression : pour cela, l'épreuve orale est incontestablement la plus sûre puisqu'elle permet toutes les nuances d'un dialogue et l'adaptation de la communication à l'interlocuteur, mais c'est aussi la moins bien acceptée.

3) Que la réponse faite soit exactement en rapport avec la question posée : un candidat peut parfaitement exposer ... comment il apprendrait la question à laquelle il ne sait pas répondre. Cela s'est vu.

En admettant que toutes les garanties nécessaires soient apportées pour que l'épreuve permette réellement de juger des moyens d'expression et que le candidat soit réellement jugé sur ce point, il reste à savoir si le but des études de médecine est de savoir faire un discours sur chacune des rubriques du catalogue, prétendu programme, ou aussi, en plus, surtout, de savoir agir. S'il est jugé impossible d'exercer la médecine avec des moyens d'expression insuffisants, qu'on en fasse une épreuve de sélection et non de compétence, car c'est justement, avant de commencer de telles études qu'on doit arriver à une maîtrise minimale de sa langue.

Quant à l'autre argument, en faveur d'une telle forme, à savoir la possibilité d'apprécier les nuances, les examinateurs eux-mêmes donnent de très fâcheux exemples si on admet qu'une "nuance" a pour fonction, non d'embellir, mais de préciser le sens du discours.

Demander:"de la corticale ou de la médullaire rénales, laquelle est la mieux vascularisée ?", dénote une absence du sens des nuances. Le candidat sage a le choix entre deux réponses : "je ne sais pas" ou "les deux sont aussi bien vascularisées". Il serait refusé,car ce qui est attendu, c'est qu'il réponde "la corticale". Il faut évidemment traduire "meilleure vascularisation" par "vascularisation plus abondante" et c'est précisément le candidat qui aurait le sens des nuances qui serait pénalisé.

Demander : "répondre brièvement" ne signifie rien. A n'importe quel sujet, on peut répondre par une phrase ou un livre. La brièveté c'est peut-être dix pages. Une phrase peut remplir une page. Une ligne peut contenir un nombre de mots très variable si les écritures ne sont pas normalisées etc... .Ou bien, le candidat répond ce qu'il veut, on ne le pénalise pas et il ne servait à rien de demander une réponse brève, ou bien il perd un temps précieux à essayer de deviner ce que peut bien représenter, dans le cas précis, la brièveté.

Demander : "justifiez votre réponse" ne signifie rien. A la limite, répondre "je dis ceci parce que le professeur l'a dit au cours" engendrerait un conflit entre deux mauvaises fois.

Enfin l'expérience a été faite de soumettre à des enseignants, tous des examinateurs, un texte à discuter, sur ce sujet précisément. Tous ces partisans de la forme discursive au nom des saintes nuances se sont divisés en trois groupes : ceux qui avaient lu le texte, ceux qui n'avaient pas tout lu, et ceux qui avaient lu ce qui n'était pas écrit. Ni dans l'expression, ni dans la compréhension(surtout à un nombre élevé d'exemplaires) des examinateurs réels ne semblent se montrer dans la réalité tels qu'ils voudraient être.

La plupart des correcteurs admettent que l'évaluation de telles épreuves est très difficile et ils ont parfaitement raison. Ce fait va très loin : un examen difficile à corriger est mal corrigé. Il n'est pas possible de lire plusieurs centaines de variations sur un thème plus ou moins bien connu sans que l'attention se relâche, sauf quelques précautions particulières :

1) Prendre un temps raisonnable, c'est à dire suffisamment long.

2) Savoir ce qu'on attend de la lecture et non commencer par s'instruire de ce qui est généralement pensé du thème d'exercice pour en faire un critère d'impressions.

3) Décomposer la réponse en autant d'éléments qu'il est nécessaire.

4) Comprendre ces éléments quels que soient l'ordre de leur présentation et la façon dont ils sont exprimés puisqu'on attache, en étant fidèle à cette méthode, de l'importance à l'expression et aux nuances.

En vérité, il semble bien, sans que cela soit démontré, que l'avantage le plus important et le moins ouvertement reconnu aux formes traditionnelles ...c'est d'être traditionnelles.

DES FORMES MODERNES

Elles ont deux prétentions : être plus en conformité avec les fins définies : une vérification objective(ce qui serait un progrès) être mieux adaptées aux conditions actuelles, ce en quoi elles seraient modernes sans pour autant se contenter d'être "à la mode". Parmi de nombreuses autres, les questions à choix multiples (QCM) sont les plus élaborées parce que les plus anciennes. Il s'agit fondamentalement d'une opération intellectuelle consistant, non à répondre à une question, mais à exprimer son choix en principe par un signe. Matériellement, la différence est considérable car les "réponses" s'expriment par des signes très simples, normalisés et peuvent être lus par une machine appropriée. Ainsi le nombre croissant de candidats n'est plus un obstacle à une correction juste, fidèle et rapide. Le "hardware" étant abondant, aucune difficulté insurmontable ne se présente dans ce domaine : la technique est adaptée à des conditions actuelles(effectifs importants sinon pléthoriques), elle est des temps actuels, elle est moderne. Ce caractère ne suffit pas pour la condamner ni pour l'adopter.

Cette technique connaît des détracteurs farouches qui fondent leur sentiment sur un fait que personne ne conteste : il en a été fait un très mauvais usage. D'autres prétendent que la part du hasard est extrêmement importante, que le bachotage suscité est pire que l'autre et encore plus stupide, sans revenir sur la querelle au sujet des moyens d'expression.

Voici une QCM, découverte dans un ouvrage belge et publiée antérieurement par le Ministère de l'Education Nationale (1961). Son origine est la Faculté de Médecine de Paris, département de Physiologie,

Soit :

1) Le débit circulatoire cutané - ET

2) La quantité de chaleur perdue par unité de temps.

A. L'augmentation du premier est accompagnée d'une augmentation du second ou la diminution du premier est accompagnée d'une diminution du second.

B. L'augmentation du premier est accompagnée d'une diminution du second ou la diminution du premier est accompagnée d'une augmentation du second.

C. Les variations du second sont indépendantes des variations du premier.

Il s'agit de choisir A, B ou C.

Voici, à peu près ce que pourrait en penser un candidat. "J'ai perdu du temps en croyant que "moins ET" représentait le débit circulatoire cutané alors qu'il fallait lire "tiret et, deuxièmement....." Il faut qu'à chaque paragraphe j'essaie de me rappeler ce que signifient "le premier" et "le second". Avec un peu d'habitude, je discerne que le premier est le premier cité deux fois en A et en B. Malheureusement, c'est le second qui est représenté comme variable indépendante en C. Si je tire au sort, j'ai une chance sur trois d"avoir" cette question très embrouillée. En fait, il y a deux traductions à faire : "premier = débit circulatoire cutané ; second = quantité de chaleur perdue par unité de temps, elle-même débit calorique".

"A" peut se résumer par "variations parallèles", "B" par "variations inverses", "C" par "indépendance". Le mot "indépendance" figure en "C". En "A" et en "B" on parle d'accompagnement et la question se pose de savoir si ce mot représente des variations liées à la même cause ou si des variations de l'un entraînent des variations de l'autre. Arrivé à ce stade des réflexions imposées par la forme même de la question, il devient logiquement soutenable de répondre soit A soit B si on suppose la température du sang constante(ce qui est typique) et la température ambiante élevée ou basse(ce qui est tout aussi logique). Pour une température ambiante basse, le débit des pertes caloriques varie avec le débit circulatoire cutané, pour une température élevée, les échanges thermiques se font en sens inverse : le débit des pertes caloriques diminue lorsque le débit circulatoire cutané augmente. La vraie question est de savoir laquelle de ces réponses était retenue comme juste.

Voici une autre QCM, tirée des mêmes sources.
Comparez X à Y
X pression mécanique dans le capillaire veineux.
Y pression oncotique " " " " .
Et dites si
A : $X > Y$
B : $Y < X$
C : $X = Y$
1) il faut comprendre que, si X est jugé supérieur à Y, il faut répondre A (convention).
2) il faut traduire "pression mécanique..." par X
3) il faut apprendre ces nouvelles conventions(inhabituelles).
4) il faut les retenir tant que l'exercice n'est pas terminé.
5) il faut faire attention dans l'expression de la réponse.

Si on juge que, dans le capillaire veineux, la pression oncotique est supérieure à la pression mécanique, il faut traduire ceci par $Y > X$ ce qui s'exprime par B !

Il aurait été aussi simple de poser la question ainsi : "comparez les deux pressions suivantes : au niveau des capillaires veineux : la pression mécanique (Pb) et la pression oncotique (Ponc) à l'aide des signes habituels : $>$, $<$, $=$ ". Pb Ponc ☐

Le correcteur accepte ou refuse la réponse : il indique son choix dans la case ménagée à cet effet. Ou, encore, en vue d'une correction automatique, on propose ainsi la réponse : Pb $>$ P.onc ☐

Pb $<$ P.onc ☐

P $=$ P.onc ☐

Tout le travail supplémentaire qu'on requiert du candidat est sans rapport avec ce qu'on veut juger : la connaissance des grandeurs relatives des pressions mécaniques et oncotiques au niveau du capillaire veineux. Ces deux connaissances sont deux objectifs parfaitement définissables.

La part du hasard peut être considérablement réduite de manière extrêmement simple. Soit une question pour laquelle cinq réponses A.B.C.D.E. sont proposées. Dans l'ignorance où on est de la question et de chacune des réponses proposées, on a une chance sur cinq de choisir la bonne à condition qu'il y en ait qu'une. S'il est dit que chacune peut être juste ou fausse, trente deux solutions sont à priori possibles et il faut reconnaître qu'il n'est pas toujours facile d'imaginer quatre réponses fausses qui aient les mêmes chances d'être reconnues comme vraies. Soit à explorer des connaissances d'un niveau relativement élémentaire sur le chien. On le définira par le nombre de ses pattes(quadrupède), par une classification élémentaire(c'est un mammifère, digitigrade) par ce qu'il mange(carnivore) et par son cri(il aboie). On vérifiera que le concept choisi du chien est bien connu lorsque ces cinq propositions seront reconnues comme vraies : l'ensemble de ces propositions constituant un item. Il n'est pas du tout nécessaire d'aligner les éléments de l'item : chacun d'eux peut se trouver dans un autre ensemble.

Ainsi :

- On fournit une liste de cinq animaux entre lesquels il faut choisir ceux qui sont quadrupèdes, parmi lesquels le chien.

- On fournit une liste de cinq animaux et de deux régimes, il faut pour chaque animal trouver le bon régime ... etc.

Finalement, on constitue, par exemple, cinq groupes de cinq propositions organisées sous un thème apparent(qu'il est d'ailleurs possible d'évaluer) mais aussi, dans chacun de ces groupes, il est possible d'isoler l'un des éléments d'un item disposé selon le schéma suivant :

1- ITEM "nombre de pattes"
A,B,chien,D,E

2- ITEM "mammifère"
chien,B,C,D,E

3- ITEM "cri" : A.chien,C.D.E.

4- ITEM "régime": A.B.C.chien,E

5- ITEM "marche": A.B.C.D.chien.

ITAM "chien" : 1C, 2A, 3B, 4D, 5E.

Un item"chien" en ligne permettrait de juger de la constance des réponses.

Une exploration aussi "serrée" qu'on le veut peut être réalisée avec une économie de propositions qui peut être considérable et des chances de répondre juste par hasard à la limite du négligeable(ici 3, 1p.100 contre 20 p.100 avec les techniques habituelles).

Quant au "bachotage", il faut s'entendre sur ce que ce mot signifie. En principe, il désigne une préparation intense à un examen ; en réalité, il voudrait plutôt suggérer que la méthode employée est stupide. Il semble bien qu'à cet égard, les QCM exposent plutôt moins à une préparation stupide que les formes traditionnelles parce que le nombre de questions possibles est beaucoup plus élevé et surtout le nombre de leurs combinaisons pratiquement infini. Soit, par exemple, à définir "saur"(comme un hareng). On peut proposer :

1. salé et séché à la fumée
2. blond
3. qui a séjourné dans la saumure
4. animal du même sous ordre que le crocodile
5. mot d'origine Viking

Trois réponses correctes sont admissibles n°s 1. 2 (1) et 5 (1) : telle que la série est constituée il n'y a aucune raison de ne pas admettre les trois. Il n'est pas non plus indispensable d'exiger les trois pour constituer un item. On peut, par exemple, refuser le choix 4 avec ce que cela comporte dans le système d'établissement des résultats, proposés au CHAP. II. Pénaliser ou non le choix 3 en fonction de la précision exigée dans le vocabulaire et accepter le choix 1 2 et 5 sans pourtant pénaliser le refus de deux de ces trois propositions.

Ainsi :

. il n'est pas accepté de prétendre que "saur" est un cousin du crocodile ou une préparation ayant séjourné dans la saumure.

. il est exigé de savoir que "saur" signifie salé et séché à la fumée.

. il n'est pas exigé de savoir que ce mot signifie "blond" en Viking mais il est mieux de le savoir.

. il ne semble pas du tout qu'une telle forme d'examen encourage particulièrement une préparation stupide à la seule condition que les examinateurs ne se fassent aucune illusion sur un système ne reposant pas sur un stock de milliers de très petites questions combinables entre elles dans une infinité de possibilités.

Ce qui semble un inconvénient réel des QCM, c'est la forme dans laquelle on enferme les candidats et qui peut être suffisamment déconcertante pour fausser les résultats. A cet inconvénients, il est possible d'opposer au moins deux solutions. La première, c'est d'entraîner les candidats à ce genre d'épreuve : on pourrait peut-être accuser cette méthode de susciter un certain bachotage et d'enfermer les candidats dans une forme très étroite. La seconde, c'est d'interposer un "noteur" entre le candidat et la grille soumise à la machine. En effet, si on pose plusieurs centaines de questions auxquelles il ne s'agit que de répondre par oui ou par non, vrai ou faux (questions fermées) ou par un mot ou par un très petit groupe de mots(questions ouvertes), le noteur peut "cocher" chaque réponse acceptée en se montrant même capable de retenir comme juste la périphrase représentant un mot oublié et dont l'oubli est pardonnable.

(1) d'après Gustave COHEN.

Un item peut parfaitement être constitué, par exemple, des propositions 1, 77, 198, 302 et 307. Contrairement à un homme, une machine convenablement programmée pourra parfaitement et sans erreur reconstituer les item. Si, en plus, chaque question est identifiée par un numéro de code déchiffrable par la machine, l'ordre de présentation des item peut être varié, ce qui diminue les chances de communication entre les candidats ou, au moins, l'efficacité de telles communications.

L'examen programmé, où le jeu consiste à parcourir une série avec un nombre minimum d'étapes, suppose une infrastructure très lourde, dont on ne sait pas très bien si elle est rentable. Il semble cependant que si, pour chaque réponse refusée, le motif du refus est immédiatement présenté, un tel type d'exercice peut apprendre quelque chose : cela semble une méthode d'enseignement riche en possibilités dont il n'est pas certain, cependant, que les avantages l'emportent sur les inconvénients d'un coût très élevé. Dans le domaine des examens, il semblerait utile que la méthode soit encore longuement testée.

Enfin, une méthode très simple et peu coûteuse de contrôle continu, mais réellement continu, peut être utilisée dans les cours suivis par un grand effectif. Il suffit d'une installation constituée d'une "machine à voter". Chaque étudiant dispose d'un pupitre constitué, par exemple, de six boutons : A.B.C.D.E. et "verrouillage". Par un moyen quelconque, le professeur propose une QCM. Chaque étudiant fait son choix et "envoie" sa réponse au moyen du "verrouillage". Un tableau totalisateur peut instantanément fournir le score de la question, score qui peut être enregistré, aussi bien que la question elle-même et le cours qui l'a précédé. La remise à zéro est commandé par le professeur. Un tel système, parfaitement réalisable, présente plusieurs avantages.

1) il permet instantanément de contrôler que l'explication est "passée" et dicte la réponse à l'enseignant qui se demandait s'il fallait ou non reprendre.

2) il permet de tester l'attention et rend stérile toute discussion sur le point de savoir si un cours d'une heure est trop long ou si, au contraire, on pourrait le prolonger.

Il mesure la durée d'efficacité d'un mode de présentation choisi, étant bien clair qu'un sujet exposé en dix minutes avec un pourcentage de réussite de 80 p. 100 est mieux exposé que si la réussite de 80 p. 100 nécessite 30 mn ou que si 10 mn de présentation sous une autre forme ne donne que 50 p. 100 de succès.

3) dans la mesure où les QCM conventionnelles sont retenues comme forme d'examen, il devient possible de les tester de façon continuelle sans aucun cérémonial particulier.

Bien entendu, un tel système est parfaitement anonyme dans la mesure où les renseignements ne sont que des totaux : c'est bien l'enseignement qui est contrôlé sans aucune sanction possible. Il serait beaucoup plus difficile, mais non irréalisable, de tester chaque étudiant, l'infrastructure nécessaire serait beaucoup plus lourde.

Au total, les moyens actuellement disponibles permettent un réel contrôle de l'enseignement permettant de motiver le choix qu'on fait de telle ou telle méthode : l'effet rétroactif de la population enseignée sur l'agent d'enseignement n'est pas utopique et c'est bien, après tout, en fonction d'une population qu'il convient d'aider, que l'aide peut être justement choisie. Un examen peut être justement et fidèlement apprécié, qu'il soit total ou partiel et même continu sans que ce dernier terme représente l'abus de langage habituel.

Mais tout ceci représente des techniques. Chacun sait qu'il n'y a pas de difficultés notables à recouvrir une pellicule d'images pendant le temps qu'on veut, mais qu'il est beaucoup plus difficile d'organiser une présentation durant un certain temps en vue d'atteindre pleinement le but choisi, ne serait-ce que distraire. Il en est exactement de même pour ces formes modernes d'examen : le support existe extrêmement maniable, il n'y a pas grand chose à mettre dessus et sur ce point, un travail considérable reste à faire afin que les immenses possibilités offertes deviennent des réalités.

NECESSITE D'EXAMEN NON VERBAUX

Pour en rester au domaine de la médecine, un fait demeure : la fonction du médecin est d'agir et non de discourir. On peut évidemment soutenir que l'acte médical, quel qu'il soit, doit être réfléchi, que la pensée doit précéder l'acte. Ceci semble incontestablement vrai dans l'exercice mais sensiblement moins sûr dans les processus d'apprentissage. C'est un fait latin certain, mais peut-être pas universel, qu'il n'y a de pensée que discursive. En ne discutant pas ce point, il faudrait donc savoir parler pour pouvoir penser ; raisonner, c'est organiser des mots en discours. Soit. Un tel objectif peut être retenu, mais ce n'est certainement pas le seul. Un pathologiste est réputé savoir discourir(dans le sens classique) sur les maladies : c'est un enseignant et un bon enseignant si son discours est compris. Ce n'est pas la fonction du médecin : il est thérapeute. Il doit être capable d'identifier les maladies. Savoir quoi prescrire dans une situation donnée, c'est un fait de connaissance. Savoir passer du malade au diagnostic suppose qu'on sait exécuter certains gestes difficiles : inspecter, palper, percuter, ausculter pour reprendre un plan bien classique. N'importe qui peut apprendre à réciter les signes d'auscultation du rétrécissement mitral. Sans prétendre que cette récitation ne prouve rien et que savoir le faire ne sert à rien, on peut avancer, en toute certitude, que cette connaissance ne sert à rien si, stéthoscope aux oreilles, on ne peut pas identifier(recueillir et interpréter) de tels signes.

Sans nier l'utilité des connaissances, il est possible d'affirmer qu'elles sont insuffisantes. Un médecin capable doit savoir passer à l'acte. Un examen doit, à un certain moment, certifier que l'étudiant est devenu capable. S'il affirme, sans vérification, une telle capacité, l'examen est un mensonge.

PERSPECTIVES D'AVENIR

Il est possible, il est temps, de rêver à ce que pourrait être dans un avenir proche, l'exercice de la médecine de tous les jours : celui du médecin praticien(qui est la règle) et non de la médecine des hôpitaux qui reste et restera l'exception.

N'importe qui pourra poser, au moins à titre d'hypothèse, les diagnostics les plus rares s'il dispose d'un terminal d'ordinateur. Dans les mêmes conditions, toutes les thérapeutiques possibles seront proposées avec posologie , contre-indications, effets secondaires...etc. Bien entendu, l'état du fond du cotyle dans une fracture de la tête du fémur, restera le résultat d'une investigation radiologique, certains traitements ne pourront être faits qu'en milieu hospitalier, les investigations nécessaires à l'établissement de certains diagnostics nécessiteront, plus que maintenant sans doute, une infrastructure tellement lourde servie par des techniciens d'une telle compétence qu'un "centre" restera nécessaire. Mais, au praticien, même s'il est moins "savant" que ses prédécesseurs(ou s'il a moins conscience de l'être) il ne restera qu'un moyen de dialoguer avec son ordinateur : lui fournir les informations qu'il aura lui-même recueillies au cours de l'examen de son malade.

Rigueur dans la recherche, rigueur dans l'expression, feront du médecin de l'ère électronique un séméiologiste d'un niveau dont il semble bien qu'on ait quelque tendance à le croire passé de mode.

CONCLUSIONS

Lorsque l'autorisation d'exercer une profession est subordonnée au succès à un examen, qui est donc une condition nécessaire et suffisante (1), c'est l'examen qui est le but et non la profession. Si on exige des futurs dentistes qu'ils sachent exécuter vingt-cinq figures de danse classique, tous les étudiants intéressés apprendront et un certain nombre finiront par savoir ce qu'on exige d'eux. Le succès à l'examen ne garantira en rien que le lauréat saura exercer son métier. Un tel succès pourrait être considéré comme une preuve de motivation, non de capacité. Garantir légalement, en attribuant un diplôme, qu'un individu est capable d'exercer une activité et d'endosser des responsabilités sans définir préalablement la valeur de l'objet garanti, c'est faire de la fausse monnaie. On sait que certains faux monnayeurs maladroits mettent en circulation des espèces dont la valeur réelle est supérieure à la valeur nominale. Inéluctablement, la tromperie, qu'elle ait pour cause l'intention de nuire ou l'insouciance, finit par tuer la confiance .

OBJECTIFS

L'enseignement médical a pour fonction de faire des médecins et il en fait. Non seulement il les fait mal mais encore il en fournit de mauvais: tous ceux qui ont fait subir des épreuves de clinique sont en accord sur ce point. Peu importe le nombre; en vérité on ne peut pas le connaître puisque chacun conçoit à sa manière ce que doit être le produit fini. Il n'y a qu'un seul remède à ce mal: définir en termes de comportements souhaitables, réalisables, observables, et vérifiables, ce que doit être le diplômé; décrire le but des études en objectifs d'examen.

(1) En Médecine, les examens ne sont pas suffisants, il faut ajouter l'inscription à l'Ordre et l'autorisation d'exercer fournie par la Préfecture.

Si on ne peut pas observer, on ne peut pas vérifier et, si on ne peut pas vérifier, on ne peut pas attribuer un diplôme, sauf si on change le sens du mot ou la valeur de l'acte. Si on ne peut pas réaliser ce qu'on souhaite, on a le choix entre deux solutions: souhaiter ce qui est réalisable ou renoncer à réaliser (et, ipso facto, fermer toutes les Facultés de Médecine).

Il ne suffit pas que les objectifs soient définis, encore faut-il qu'ils soient décrits en termes convenables et que cesse d'avoir cours cette conception démentielle, qu'il suffit de décrire un acte pour savoir le faire. Il est absolument nécessaire que tous les comportements souhaités ne soient pas du seul domaine verbal pour des gens qui ont à concevoir des actes et à les exécuter. Si le seul comportement, reconnu utile pour un médecin, est de savoir faire des discours sur un nombre, même élevé de maladies, il ne faut pas attester qu'il est capable de les reconnaître (et de les traîter) mais simplement qu'il sait en parler. Peut être est-il en plus et comme par hasard, capable de les reconnaître? Si on n'en a pas la preuve, on ne peut pas le garantir .

Au total, la définition des objectifs permet de répondre à deux questions: quel but faut-il atteindre? C'est l'objectif des études . Comment savoir si le but est atteint? C'est l'objectif des examens . Savoir répondre à la première seulement, permet de rêver, pas d'agir.

RESULTATS

En admettant que le résultat d'un examen soit la conséquence d'une opération analogue à une mesure, les critères d'un bon résultat sont ceux d'une bonne mesure.

Fidélité et sensibilité sont deux qualités attachées à la méthode précédemment et longuement décrite et qui semble originale.
La justesse n'est que relative dans la mesure où un étalon n'est pas défini. La définition des objectifs en termes d'examen permet en outre, d'une part, d'assurer la sélectivité de l'opération (on sait ce qu'on mesure) et, d'autre part, d'affirmer que la chose mesurée

est en rapport certain avec la nature même du but des études, si la définition des objectifs est celle qui convient. On peut affirmer que le but est atteint si le candidat est parfait. On ne peut pas savoir d'un candidat qu'il est satisfaisant s'il est imparfait, sauf définition d'une valeur relative qui ne peut être, dans l'état actuel des choses, qu'empirique ou arbitraire. Cependant, étant admise l'imperfection de la définition du niveau satisfaisant, la situation du candidat, par rapport à cette référence, peut s'évaluer avec une précision significative aussi grande qu'on le souhaite.

Au total, il ne semblerait pas déraisonnable que le diplôme (examen réellement terminal) , attestant que le diplômé possède les capacités qu'il prétend, soit constitué de telle sorte que tout candidat reçu ait parfaitement réussi l'épreuve qui prouverait ainsi, non un niveau de valeur personnelle, mais un niveau minimal: le minimum obligatoire. La conception d'un tel examen serait certainement laborieuse, mais son aboutissement heureux serait de fournir une garantie aussi sérieuse que le seraient les examinateurs eux-mêmes. Avant de rejeter l'idée d'une épreuve n'exigeant qu'un minimum, il conviendrait de se convaincre que, dans ce domaine, un maximum reste utopique et une moyenne indéfinissable. Il conviendrait encore de méditer sur la garantie réelle qu'offrent des examens dont, seul, est élevé le niveau des prétentions .

PROGRAMME

Dans des études nécessairement longues, il est nécessaire de progresser.

De même que le travail fourni est un moyen et non le but, que le nombre de litres de sueur fournis par le front laborieux n'atteste en rien la qualité du produit fini, qu'un diplôme ne doit pas être la récompense d'un courage, même exemplaire, mais la garantie de capacités définies, c'est la progression qui doit définir un programme et non la durée du parcours, insuffisante pour préjuger de l'atteinte du but. Si la fonction des enseignants est d'aider les étudiants dans leur progression et non d'exécuter des numéros de saltimbanques ,

ce qui permet de juger de l'utilité de leur action, ce n'est pas la rumeur admirative du peuple mais leur efficacité, exprimée en termes de progression, de la population dont ils ont la responsabilité.

Si la méthode proposée d'établissement des résultats ne permet pas, à elle seule, de définir le but à atteindre, elle permet, non seulement de situer chaque candidat par rapport au but choisi, mais encore de comparer, avec une grande précision, une situation actuelle avec une situation précédente et, par conséquent, de mesurer une progression, si on utilise le coefficient de réussite.

L'autonomie des universités, inscrite dans la loi, peut s'inscrire dans les faits. Si le but est commun pour toutes les Facultés de Médecine et si ce but est décrit, les moyens choisis pour y parvenir peuvent être différents dans chacune d'entre elles. Méthodes et programmes peuvent parfaitement être définis en termes différents. Le danger serait extrême si le critère de choix était la séduction. Le risque propre à la disparité des moyens est nul si le but est commun, la progression contrôlable et contrôlée . Ceci suppose que le " service des examens" soit un service d'enseignement et non un service administratif ayant pour seule fonction d'enregistrer des résultats .

L'aboutissement de cette réflexion, c'est qu'il reste beaucoup à faire. Une méthode de critique est disponible et le critère du travail à faire c'est l'efficacité: c'est en fonction de la réussite entraînée que pourront être retenus tels programmes ou telles méthodes. Avant de choisir les moyens, il faut évidemment recenser ceux qui existent, éventuellement imaginer ceux qui manqueraient. Mais, surtout, avant de savoir "comment faire", il faut savoir "quoi faire". La définition des objectifs est un impératif catégorique et prioritaire. C'est un travail gigantesque mais ... la difficulté de réussir ne fait qu'ajouter à la nécessité d'entreprendre.

TABLE DES MATIERES

PREMIERE PARTIE : UN MAL

I : DES EFFECTIFS 2
II : DES NOTES 32
III : DES NOTES ET DES EFFECTIFS 57
IV : GRANDEURS D'UN EXAMEN 69
V : GRANDEURS COMPARABLES ? 83
VI : UN JURY ET DES NOTES 107
VII : LE MAL 131

DEUXIEME PARTIE : UN REMEDE

A — POSSEDER UN VOCABULAIRE PRECIS

I : LES MOTS 149

B — DONNER UN SENS AUX RESULTATS ET UNE SIGNIFICATION AUX NOTES

II : ETABLISSEMENT DU RESULTAT 181
III : DU RESULTAT A LA NOTE 214
IV : UNE NOTATION NORMALISEE 243

C — DEFINIR LES OBJECTIFS

V : UTILITE D'UN EXAMEN 265
VI : OBJECTIFS D'UN EXAMEN 271
VII : EXAMEN ET CONTROLE DES CONNAISSANCES 283
VIII : OBJECTIFS DEFINIS D'UN EXAMEN 297

D — ETABLIR UN PROGRAMME

IX : PROGRAMMES 307
X : CIRCONSTANCES D'UN EXAMEN 330

E — CHOISIR UNE FORME ADEQUATE

XI : FORMES D'UN EXAMEN 352

CONCLUSIONS 372

BIBLIOGRAPHIE

1- G. BARRES : Une nouvelle méthode de notation du contrôle des connaissances, communication personnelle, 1971.

2- B.S. BLOOM et coll.: Taxonomy of educational objectives, David McKay, 1956.

3- J. GOSSET : Eléments de pédagogie médicale, l'Expansion, 1969.

4- M. LAMOTTE : Initiation aux méthodes statistiques en biologie Masson, 1967.

5- G. de LANDSHEERE : Evaluation continue et examens. Précis de docimologie, Labor-Nathan, 1971.

6- R. MAGER : Preparing Instructional Objectives, Fearon, 1962.

7- H. PIERON : Examens et Docimologie, P.U.F., 1969.

8- P. ROUSSEAU : Histoire de la Science, A. FAYARD, 1945.

9- D. SCHWARTZ : Méthodes statistiques à l'usage des médecins et des biologistes, FLAMMARION, 1963.

-:-:-:-:-:-:-:-

ANNEXES

I

CHIRURGIE GENERALE

Arrêté du 25 Avril 1961

Art. 7 (modifié par l'arrêté du 7 Septembre 1967). — L'examen de fin d'études comporte les épreuves suivantes:

1/ Deux épreuves écrites anonymes portant respectivement sur la pathologie chirurgicale et sur la thérapeutique chirurgicale (durée globale: trois heures);

2/ Une épreuve clinique consistant en un examen de malade (durée: une heure).

Chacune des épreuves est notée de 0 à 20. Nul ne peut se présenter à l'épreuve clinique s'il n'a obtenu au moins 20 points aux épreuves écrites. Sont déclarés admis les candidats ayant obtenu pour l'ensemble des épreuves écrites et clinique une note moyenne au moins égale à 10 sur 20.

II

GYNECOLOGIE MEDICALE ET OBSTETRIQUE

Arrêté du 14 Janvier 1955

(Modifié par arrêté du 1er Juin 1955)

Art. 13 (modifié par l'arrêté du 5 juin 1964). — Pour le certificat d'études spéciales d'obstétrique, l'enseignement est sanctionné par un examen de fins d'études.

Il y a une session d'examen par an. L'examen comporte:

a) Des épreuves écrites anonymes:
Une épreuve portant sur l'anatomie de l'appareil génital féminin ou sur la physiologie gravidique ou sur ces deux matières (durée: deux heures; notée de 0 à 20).
Une épreuve portant sur l'obstétrique ou sur le nouveau-né ou sur l'obstétrique et le nouveau-né (durée : deux heures ; notée de 0 à 30).

Toute note inférieure à 5 est éliminatoire. Pour être admis à subir les épreuves cliniques et techniques, les candidats doivent obtenir un nombre de points au moins égal à 25 aux épreuves écrites;

b) Des épreuves cliniques et techniques:

1/ Une épreuve clinique comportant l'examen en quinze minutes d'une femme enceinte ou d'une parturiente, ou d'une accouchée, ou d'un nouveau-né, suivi, après dix minutes de réflexion, d'un exposé oral de dix minutes, puis d'une interrogation portant sur une ou plusieurs des matières suivantes: pathologie de la femme enceinte, de l'accouchement et des suites de couches, l'allaitement et ses troubles, les interventions obstétricales, l'obstétricie sociale, la sérologie, la radiologie en obstétrique;
2/ Une épreuve de thérapeutique.

Chaque épreuve est notée de 0 à 20. Pour être admis à l'examen, les candidats doivent obtenir un nombre de points au moins égal à 20 pour les épreuves cliniques et techniques.

III

ELECTRORADIOLOGIE MEDICALE

Arrêté du 11 Octobre 1948.

(Modifié par les arrêtés des 22 juin 1949, 13 octobre 1949, 29 novembre 1950, 4 septembre 1951, 26 octobre 1951, 15 juillet 1952 et 17 juillet 1953).

L'examen de fin d'études a lieu chaque année au mois d'octobre. Il comporte deux parties:

Première partie
(Modifié par l'arrêté du 19 octobre 1964).

1/ Une épreuve écrite anonyme portant sur un sujet de radio-diagnostic (0 à 30);

2/ Une épreuve écrite anonyme portant sur un sujet de roentgenthérapie ou de curiethérapie (0 à 30);
3/ Une épreuve écrite anonyme portant sur un sujet d'électrophysiothérapie (0 à 20).
Les épreuves écrites ont une durée globale de trois heures et demi (trois heures de rédaction après une demi-heure de réflexion).
Pour être admis à subir les épreuves de la deuxième partie, les candidats doivent obtenir un nombre de points au moins égal à 40 pour les épreuves de la première partie.

Deuxième partie

1/ Composition et discussion d'un mémoire écrit, rédigé en cinq exemplaires (0 à 20);

2/ Une épreuve clinique : examen d'un malade, prise ou lecture de clichés (0 à 20);

3/ Une épreuve orale portant sur l'ensemble du programme (0 à 20).

Pour être admis à l'examen de fin d'études, les candidats doivent obtenir au moins la moitié du maximum des points pour chacune des épreuves de la deuxième partie (arrêté du 17 juillet 1953).
Cependant, toute note inférieure à cinq à l'une des épreuves de la première ou de la deuxième partie est éliminatoire.

VU, LE PRESIDENT DE THESE

A. MONSAINGEON

VU, LE DOYEN DE LA FACULTE

A. MONSAINGEON

VU, LE PERMIS D'IMPRIMER

LE RECTEUR DE L'ACADEMIE DE PARIS

R. MALLET